APPLICATIONS OF THE HELMHOLTZ-HODGE DECOMPOSITION TO NETWORKS AND RANDOM PROCESSES

by

ALEXANDER STRANG

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy

Dissertation Advisor: Dr. Peter Thomas

Department of Mathematics, Applied Mathematics and Statistics

CASE WESTERN RESERVE UNIVERSITY

August 2020

CASE WESTERN RESERVE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

We hereby approve the dissertation of

Alexander Strang

candidate for the Doctor of Philosophy degree *

Committee Chair: Peter Thomas
Dr. Peter Thomas
Dissertation Advisor
Professor,
Department of Mathematics, Applied Mathematics and Statistics

Committee: Daniela Calvetti
Dr. Daniela Calvetti
Professor,
Department of Mathematics, Applied Mathematics and Statistics

Committee: Wojbor Woyczynski
Dr. Wojbor Woyczynski
Professor,
Department of Mathematics, Applied Mathematics and Statistics

Committee: Karen Abbott
Dr. Karen Abbott
Professor,
Department of Biology

Committee: Michael Hinczewski
Dr. Michael Hinczewski
Professor,
Department of Physics

August 13, 2020

*We also certify that written approval has been obtained for any proprietary material contained therein.

Table of Contents

List of Tables

List of Figures

Acknowledgement

This thesis project is the result of four years of collaboration. I am deeply grateful to everyone who has supported me throughout the past four years. I would not have come this far without you.

To Peter Thomas - thank you for encouraging my intellectual curiosity, and guiding it towards interesting questions. It has been a joy working with you. I will fondly remember the excitement of talking through ideas and developing new questions together. To Karen Abbott - thank you for your enthusiasm and your support. It was wonderful to be welcomed into your lab. Working collaboratively with the other students in the lab was a highlight of my time in the graduate program. Thank you both for your time, your generosity, and your faith in me.

To Daniela Calvetti - thank you for all of your mentorship. You brought me into the graduate program here, and your advice has been invaluable. It was your class on linear algebra that motivated me to pursue applied math as an undergraduate. Thank you for teaching me to pursue questions with rigor and diligence, and for all your help finding new professional opportunities.

To Michael Hinczewski and Robin Synder - thank you for always asking challenging questions. Our conversations together never failed to sharpen my work. Working with you helped me think about problems from different perspectives and encouraged me to think about applications in different fields.

I am grateful to all of my committee members for their time, insight, and expertise. I

am also grateful to all of the professors who helped me grow as a student. I am especially grateful to Wojbor Woyczynski, whose lectures were as unerringly entertaining as they were enlightening, to Elizabeth Meckes, whose classes pushed me to be a better mathematician, and to Erkki Somersalo for his clarity, patience, and kindness. I would also like to thank Donyear Thomas and the department staff for all of their essential help behind the scenes.

The work presented in this thesis is inspired by Lek-Heng Lim's work on the Helmholtz-Hodge Decomposition. I am grateful to him for the conversations we had four years ago that started me down this path. Daizaburo Shizuka and Michel Regenwetter also deserve mention for generously sharing some of the data analyzed in this thesis.

To my mother and father - thank you for everything. To my grandfather - I am lucky to have had you behind me all this way. You are an inspiration. Hopefully there is some linear algebra here that you will enjoy.

Applications of the Helmholtz-Hodge Decomposition to Networks and Random Processes

Abstract

by

ALEXANDER STRANG

Cycle and potential decompositions are widely used mathematical tools for analyzing systems across fields. The discrete Hodge-Helmholtz Decomposition (HHD) decomposes an edge flow on a graph into two components. The first component is conservative, and associated with the gradient of a potential function defined on the vertices. The second component is cyclic, and is associated with the adjoint curl of a set of vorticities defined on the loops of the network. We explore applications of the HHD to problems arising in a variety of fields. We provide examples where the HHD is used as a descriptive tool for characterizing structure, and as a predictive tool for understanding dynamics. To show that the HHD can be used to describe the structure we apply it to tournaments arising in politics, animal behavior, and sports. To show that the HHD can be used to predict dynamics we apply the decomposition to discrete-space continuous-time Markov models motivated by biophysical and ecological examples. It is shown that the HHD has a natural thermodynamic interpretation, and can be used to construct analogous thermodynamics for a generic class of Markov chains. We show that the HHD can be applied to understand steady-state dynamics in either the strong noise, or weak rotation, limit and controls the

long-term production rate of observables. A formal expansion of steady-state distributions and steady-state fluxes in the cyclic component of the HHD is introduced. Comparisons to existing potential theories and cycle decompositions are made, and it is shown that the HHD is a complementary decomposition to the quasipotential in the continuum limit.

Part I

Introduction

Chapter 1

Introduction

1.1 Outline

Potentials and cycle decompositions play a critical role in the analysis of many physical systems. Problems in areas as diverse as physics, chemistry, and biology can often be simplified by considering underlying potentials and tendencies to cycle. Potentials are a classical topic with deep roots in physics [1, 2]. Cycle decompositions, in particular, cycle decompositions on networks, are a topic of active interest in the analysis of biochemical systems and in biophysics, since many cellular functions rely on molecular machinery that can cycle through a set of states in order to complete a task (cf. [3, 4, 5, 6, 7]). Cycles are of broad interest in biology in general, as cyclic are important at both the microscopic and macroscopic levels.

This dissertation will explore applications of the Helmholtz-Hodge Decomposition (HHD). The HHD separates a flow into a conservative component associated with the gradient of a potential function, and a cyclic component associated with the adjoint curl of a set of

vorticities. The HHD generalizes the familiar Helmholtz decomposition widely used in electromagnetism and fluid mechanics [2, 8]. The Helmholtz decomposition has broad applications including computer graphics, flow visualization, astrophysics, imaging, and robotics [9, 10, 11, 12, 13, 14]. A survey of applications is available in [8]. Notably, none of the applications addressed in the survey are stochastic, or are based on a discrete state space described by a network.

The combinatorial/discrete HHD generalizes the Helmholtz decomposition to edge flows on networks [15, 16]. Our work is largely inspired by the Lek-Heng Lim's work on the combinatorial HHD (cf. [15] and [16, 17]). In this dissertation we illustrate the usefulness of the HHD as both a descriptive and predictive analytic tool. The HHD can be used to characterize structural properties of an edge flow on a network, and thus to describe the flow. If the edge flow is chosen to reflect the dynamics of a process developing on the network, then the HHD can be used to predict dynamics. The HHD has been used to solve ranking problems that arise in data science [15, 16, 17], and to study optimal strategies in games that involve cyclic structures [18]. We aim to illustrate the power of the HHD in different settings. To show its descriptive value we apply the HHD to study the structure of competition in tournaments. To show its predictive value we apply the HHD to Markov processes on networks.

The dissertation is organized as follows. First, in Chapter 2, we define the decomposition and review techniques for performing the decomposition. Basic properties of the operators involved are discussed and special attention is devoted to planar networks in order to highlight symmetries in the decomposition. The decomposition is characterized using least squares, discrete Poission equations, and a novel "path integral" interpretation that represents potential differences as the average work to move between two points over randomly drawn paths. Two important generalizations to the HHD are then introduced.

Solution methods for select networks are highlighted. In Chapter 3 explicit methods for certain simple network topologies are provided. In order to understand the decomposition when applied to square lattices we analyze the decomposition under graph products. The chapter concludes with implicit numerical techniques for generic networks. These techniques are designed to leverage sparsity for efficiency since application networks may be very large.

Part II considers the application of the HHD to describe structure. Chapter 4 and Chapter 5, consider the application of the HHD to competitive tournaments. These chapters use the HHD to describe the structure of competition, and are based on a paper submitted to SIAM Review and a following paper that is in preparation. Consequently, Chapters 4 and 5 are self-enclosed, and could be read independently.

Chapter 4 motivates the use of the HHD, and provides an interpretation of the components, in the context of tournaments. This interpretation is essential since the descriptive utility of the decomposition depends primarily on its interpretability. Chapter 4 concludes by analyzing the expected structure of tournaments drawn from different null distributions. Special attention is devoted to the class of trait-performance models, in which the probability one competitor beats another can be expressed as a function of their traits, which are sampled from a trait distribution. By focusing on trait-performance models we can rigorously define and prove heuristic statements about cyclic competition that appear throughout the literature. Chapter 5 develops an estimation framework for applying the HHD to data drawn from real-world tournaments. Bayesian point estimation, posterior sampling, interval estimation, and hypothesis testing are all discussed and details are provided in the Appendix (see Appendix A, Appendix B, and Appendix C). The methods are then applied to examples from politics, animal behavior, and sports. The political examples include analysis of the 2016 and 2020 presidential elections. The interpretation and significance of

the decomposition for each example area is discussed, and the structure of competition is compared across examples.

Part III considers the use of the HHD to analyze and predict dynamics of random processes on graphs. Chapter 6 introduces Markov chains on networks, and develops the rudiments of the decomposition when applied to random walks. Special attention is paid to the detailed balance case. A physical interpretation of the components of the HHD is developed. The physical interpretation closely mirrors Schnakenberg's formulation of nonequilibrium statistical mechanics [5, 19], and the axiomatic thermodynamics proposed by Qian [20, 21, 4, 22].

Analysis of steady-state, steady-state fluxes, and observable production is continued in Chapter 7. In Chapter 7 we show that, for any discrete-space Markov Chain satisfying microscopic reversibility, the problem of finding a nonequilibrium steady-state can always be transformed into the problem of finding a nonequilibrium steady-state for a process that is purely cyclic. Thus the generic steady-state of nonequilibrium processes can be understood by understanding the steady-state of purely cyclic processes. It is then shown that, in the weak rotation limit, the steady-state distribution, steady-state fluxes, and long term production rate of observables can all be expanded using an recursive sequence of Helmholtz-Hodge Decompositions. Therefore the decomposition is fundamental to the dynamics of nonequilibrium Markov processes that are close to satisfying detailed balance (close to an equilibrium process). We also present a limiting analysis of nonequilibrium steady states for processes that are dominated by diffusion, and process that are dominated by drift. It is shown that in the former case that the HHD is the appropriate cycle space decomposition for understanding the steady state, and the steady state is asymptotically described by the associated potential. In a strong drift limit a different potential is needed. This potential mimics the quasipotential [23] that is widely used to analyze steady states of

stochastic differential equations (SDE) in a small noise limit.

In Chapter 8 we analyze the convergence of the decomposition in a system size limit. It is shown that, under regularity conditions on the network, in a continuum limit the HHD converges to a Helmholtz decomposition of a vector field associated with the diffusive process achieved as a limit of the discrete space process. The associated equations are compared to the equations defining the Friedlin-Wentzell quasipotential [23, 24, 25] which is widely used to analyze steady-states and first-passage times in the weak noise limit. The path integral interpretation of the HHD is then compared to the WKB approximation. It is shown that the steady-state distribution of a generic Stochastic Differential Equation (SDE) is a PDE whose operator is a weighted combination of the operators defining the Helmholtz potential and the quasipotential. When noise is weak the quasipotential term dominates, but when noise is strong the Helmholtz term dominates. Thus the potential associated with the HHD is most relevant to diffusive processes in the strong noise limit.

We discuss our results and possible future work in Chapter 9.

Part II

The Decomposition

Chapter 2

The Decomposition

2.1 Preface

This chapter introduces the discrete Helmholtz-Hodge decomposition (HHD). The rest of the dissertation is devoted to applications of the HHD. The decomposition is defined (Section 2.2), properties of the operators and subspaces associated with the decomposition are discussed, solution methods are presented, and different interpretations are proposed (Section 2.3). The chapter concludes by introducing two generalizations that are useful when applying the HHD to certain networks, and when studying the relationship between the HHD and the dynamics of a Markov process (Section 2.4).

2.2 The Helmholtz-Hodge Decomposition

2.2.1 The Decomposition in the Continuum: A Brief Review

Since the proposed decomposition is inspired by potentials on conservative vector fields it is natural to start with the familiar formalism. This section is used to review the basic features of a conservative vector fields and the Helmholtz decomposition [2]. These properties can then be generalized to networks.

Suppose that for every point $x \in \Omega \subseteq \mathbb{R}^3$ there exists a vector $v(x) \in \mathbb{R}^3$. Then $v(x)$ is a vector field $v : \Omega \to \mathbb{R}^3$. We will assume that the domain Ω is simply connected.

Now consider a continuous path $y(t)$ through $\mathbb{R}^3$ from point a to point b. Here the path is parametrized in t, however no parametrization is necessary. The path integral of $v(x)$ over $y(t)$ is defined:

$$I(y|v) = \int_a^b v \cdot dy.$$

The path integral represents how much work the vector field does along the path.

A vector field is conservative if the path integral over any closed cycle is zero. If $v(x)$ is conservative then any path integral depends exclusively on the endpoints, not the path taken. That is, the path integral is path independent.

These two properties are the heart of conservative vector fields. The first requires that $v(x)$ has no tendency to circulate. The second requires that the work done by $v(x)$ over a path depends exclusively on the end points of the path.

It follows that there must exist a scalar function $\phi(x)$ such that:

$$I(y|v) = I(a, b|v) = \int_a^b v \cdot dy = \phi(a) - \phi(b)$$

and:

$$v(x) = -\nabla\phi(x). \tag{2.1}$$

Equation (2.1) is the standard connection between a scalar potential ϕ and a conservative vector field v. The vector field points along the direction of steepest descent of the potential. Local maxima in the potential correspond to sources in the vector field, and local minima correspond to sinks. Consider the motion of a system driven by the vector field v. If $x(t)$ represents the state of the system at time t, then $\frac{d}{dt}x(t) = v(x) = -\nabla\phi(x)$. In this context x flows downhill along the potential. It is clear that equilibria correspond to extrema in the potential since the system stops moving where $\nabla\phi(x) = 0$. Stable equilibria correspond to local minima, and unstable equilibria correspond to local maxima or saddle points.

Equation (2.1) lays the groundwork for finding a network potential. Where there is an analogy to a path integral which is zero over all closed cycles, we can apply the same formalism to find a scalar potential whose gradient recovers the original field.

The Helmholtz-Hodge Decomposition generalizes this picture by decomposing arbitrary vector fields into the ranges of two differential operators. The first differential operator is the gradient, and is associated with the scalar potential ϕ. The second differential operator is the curl, and is associated with the vector potential [8]. Depending on the boundary conditions an additional harmonic vector field may also be needed which describes translation. The curl of a vector field, denoted $\nabla \times v(x)$, expresses the tendency of v to circulate at x. Formally the curl maps each point in a vector field to a vector that represents the circulation of the original field. The magnitude of the output vector describes how much the field circulates; the direction of the output vector describes the direction of circulation.

The curl can be defined as the limit of path integrals over closed cycles. Consider a path y that forms a closed cycle $\mathcal{C}$ in a plane oriented normal to the direction $\hat{n}$ about the point x containing an area A. Then the curl is the value of the path integral in the limit that its area goes to zero:

$$(\nabla \times v(x)) \cdot \hat{n} = \lim_{A \to 0} \frac{1}{A} \oint_{\mathcal{C}} v \cdot dy. \tag{2.2}$$

If a vector field is conservative then the curl is zero everywhere. Conveniently the converse is also true under appropriate assumptions on the domain. If the curl of a vector field is zero everywhere, and Ω is simply connected, then the vector field is conservative.

The proof follows from Stokes' Theorem. Stokes' Theorem relates the path integral over a closed cycle to the total curl of the vector field inside the cycle. Loosely, the path integral over a cycle is the integral over the curl inside the cycle. Formally:

$$\oint_{\mathcal{C}} v \cdot dy = \iint_{\mathcal{S}} (\nabla \times v(x)) \cdot \hat{n} dx$$

where the path integral is taken over loop $\mathcal{C}$, which contains the oriented surface $\mathcal{S}$ with orientation $\hat{n}$. The orientation of the surface is chosen to match the orientation of the path integral. Here $(\nabla \times v(x)) \cdot \hat{n}$ is the component of the curl at x in the orientation of the path integral. If the cycle is precessed clockwise then $(\nabla \times v(x)) \cdot \hat{n}$ is the clockwise component of the curl minus the counterclockwise component of the curl.

Now suppose the the curl is zero everywhere. Then the path integral around any closed cycle is also zero:

$$\oint_{\mathcal{C}} v \cdot dy = \iint_{\mathcal{S}} (\nabla \times v(x)) \cdot \hat{n} dx = \iint_{\mathcal{S}} 0 dx = 0.$$

It follows that a vector field is conservative if the curl of the vector field is zero every-

where. That is, irrotational vector fields are conservative if the domain is simply connected. A guide to terminology is in Table 2.1. Moreover, since $\nabla \times \nabla\phi(x) = 0$ for any $\phi(x)$, any conservative vector field is curl-free. Therefore:

Lemma 1 (Conservative Vector Fields)**.** *If $\Omega \subset \mathbb{R}^3$ is a simply-connected domain then a continuously differentiable vector field $v(x)$ on Ω is conservative if and only if $\nabla \times v(x) = 0$ for all $x \in \Omega$.*

Lemma 1 sets the stage for a more general discussion of the desired decomposition.

Suppose we are given two linear operators A and B. For continuous vector fields these may be thought of as differential operators, and for networks may be thought of as matrices. The nullspace of an operator is defined as the space of objects that return zero when acted on by the operator. Lemma 1 states that conservative vector fields live in the nullspace of the curl. The range of an operator is the space of possible outputs of the operator. Since any conservative field is the gradient of a scalar potential it follows that any conservative field lives in the range of the gradient. Therefore the range of the gradient is contained in the nullspace of the curl.

If B is associated with the gradient, and A with the curl, then:

$$AB = 0. \tag{2.3}$$

since the range of the gradient is in the nullspace of the curl. This orthogonality is the heart of the Hodge decomposition [15], and is a weaker form of Lemma 1.

From the fundamental theorem of linear algebra, any finite dimensional vector space V can be decomposed into the nullspace of a linear operator, and the range of its conjugate transpose [26]. A stricter equivalent statement for vector fields is the fundamental

theorem of vector calculus, or Helmholtz's theorem, which states that any continuously differentiable vector field in $\mathbb{R}^3$ that decays sufficiently fast can be expressed as the sum of a conservative vector field (range of gradient) and an incompressible vector field (nullspace of divergence) [2, 27]. Helhomtz's theorem is a special case of the Hodge decomposition theorem [27]. That is:

$$V = \text{null}\{A\} \oplus \text{range}\{A^*\}. \tag{2.4}$$

Since $\text{range}\{B\} \in \text{null}\{A\}$ we can decompose $\text{null}\{A\}$ into its intersection with the range of B and the nullspace of B^*:

$$V = \text{range}\{B\} \oplus (\text{null}\{B^*\} \cap \text{null}\{A\}) \oplus \text{range}\{A^*\}. \tag{2.5}$$

Equation (2.5) defines the Helmholtz Hodge Decomposition (HHD) [15] for an arbitrary pair of linear operators satisfying Equation (2.3). If we associate B with the gradient then the range of B is the space of conservative vector fields. If we associate A with the curl then the range of A^* is the space of rotational (solenoidal) vector fields. The middle term represents the shared null space of the two operators. The Hodge Laplacian is defined $A^*A + BB^*$. The shared null space $(\text{null}\{B^*\} \cap \text{null}\{A\})$ is equivalent to the nullspace of the Hodge Laplacian. Vector fields in the null space of the Hodge Laplacian are harmonic (the value of the field at an interior node is equal to an average of the field over the neighboring nodes) [8, 15].

So, given any vector field v we can write:

$$v(x) = v_{\text{con}}(x) + v_{\text{rot}}(x) + h(x) = -\nabla\phi(x) + \nabla \times A(x) + h(x) \tag{2.6}$$

for some scalar potential $\phi(x)$, vector potential $A(x)$ and harmonic vector field $h(x)$ [8].

The range of the gradient $v_{\text{con}}(x)$ is the component of the vector field that tends to diverge away from sources and converge towards sinks. This is the conservative field. The range of the adjoint curl $v_{\text{rot}}(x)$ is the component of $v(x)$ that tends to rotate. This is the rotational, or solenoidal, field. It is used to describe incompressible fluid flow and the magnetic field in electromagnetism. Since the conservative field is in the nullspace of the curl it is sometimes referred to as the irrotational field.

Names	Meaning	Properties
Irrotational	null$\{\nabla\times\}$	Curl Free, Path Independent
Incompressible, Solenoidal	null$\{\nabla\cdot\}$	Divergence Free, Dynamic Equilibrium
Conservative	range$\{\nabla\}$	Reversible, Static Equilibrium, Detailed Balance
Rotational	range$\{\nabla\times\}$	Divergence Free

Table 2.1: Terminology from different fields describing the range and nullspace of the differential operators.

The harmonic field $h(x)$ is the component of $v(x)$ that neither converges, nor rotates. Since $h(x)$ lives in the nullspace of two differential operators we can think of $h(x)$ as a constant background flow that corresponds to translation [2]. Vector fields in $h(x)$ neither originate from sources, or flow into any sinks. They also never form any closed cycles. As a result, if the boundaries of the vector field are closed (no flow out of or into the boundary) then $h(x)$ must be zero. This follows from the maximum principle for harmonic functions. A harmonic function can only achieve its maximum on its boundary [28]. So, if the flow through the boundary is zero then $h(x)$ must be zero everywhere. More subtly, if the vector field is infinite, and the flows converge to zero faster than the surface area of the boundary of the domain diverges (as we expand the boundaries towards infinity) then $h(x)$ is also zero [29].

Therefore, at least for domains with closed boundaries, there exists a unique decomposition of $v(x)$ into a conservative and rotational field [29, 8].

To prove uniqueness we must show that the nullspace of the curl is the range of the gradient for vector fields on a closed domain. From Lemma 1 all conservative vector fields live in the nullspace of the curl. If a vector field is conservative (in the range of the gradient), then it is impossible for any path integral to be nonzero since it is impossible to move in a closed cycle and end higher or lower than one started on the scalar potential. This is not necessarily true if the domain has open boundaries or is infinite [30].

Lemma 2 (Uniqueness for Closed and Bounded Domains)**.** *If the domain Ω containing vector field $v(x)$ is bounded, the boundary of the domain is closed, and $v(x)$ is twice continuously differentiable, then there exists a unique decomposition:*

$$v(x) = v_{\text{con}}(x) + v_{\text{rot}}(x) = -\nabla\phi(x) + \nabla \times A(x). \tag{2.7}$$

Lemma 2 is a weaker form of a more uniqueness general theorem which specifies the flow through the boundary of an open domain [31].

We will mostly restrict our discussion to the case when $h(x) = 0$, since for any finite network the harmonic component is necessarily zero. A more general discussion including treatment of $h(x)$ is included in the section on the open boundary problem (see Section 2.4.3).

This completes the basics of the Helmholtz-Hodge Decomposition (HHD) needed to build an analogous theory on networks. Before moving on it is worth taking a moment to review the operators involved, and their significance.

1. The gradient, ∇, computes the direction and magnitude of the fastest route of ascent along the surface of a scalar function. In $\mathbb{R}^n$ $\nabla = [\partial_1, \partial_2, ...\partial_n]^\intercal$. Since the analogous network operators are matrices, we will let G represent the network gradient. Conservative vector fields live in the range of G and can be written as the gradient of an

underlying scalar potential. The gradient maps from scalar functions to vector fields.

2. The divergence $\nabla\cdot$, is the adjoint of the gradient. The divergence measures the tendency of a vector field to diverge from a point. This can be expressed as the transpose of the gradient, or, the limit of the net-flux out of a bounded domain as the domain shrinks to zero. The analogous discrete operator is denoted $D = -G^\intercal$. The divergence of a rotational field is always zero. The divergence maps from vector fields to scalar functions, and obeys the divergence theorem.

3. The curl $\nabla\times$, measures the rotation of a vector field at a single point. In continuous space the curl maps from the space of vector fields on $\mathbb{R}^3$ to the space of vector fields on $\mathbb{R}^3$. In the discrete case, the matrix C represents the curl. Conservative vector fields live in the nullspace of the curl. The curl operator obeys Stokes' theorem.

4. The adjoint curl $\nabla^*\times$, is the conjugate transpose of the curl operator. The discrete adjoint curl is $C^\intercal$. In the continuous case, the adjoint curl maps from the space of vector fields on $\mathbb{R}^3$ to the space of vector fields on $\mathbb{R}^3$ and is equivalent to the curl. In the discrete case, the curl and adjoint curl will not map to the same spaces, so we will be careful to distinguish them.

Finally, the operators are orthogonal:

$$\nabla\cdot(\nabla^*\times) = 0 = \nabla\times(\nabla). \tag{2.8}$$

In order to define a meaningful network HHD we need to find matrices G and C such that Equation (2.8) is satisfied ($G^\intercal C^\intercal = 0 = CG$), and the operators retain their qualitative significance. In addition we would like the discrete operators to retain as many of the properties of the differential operators as possible. In Section 2.2.3 we will check that

G and C retain the appropriate product rules, and that both the divergence theorem and Stokes' theorem still apply.

The stage is now set to generalize the HHD to networks.

2.2.2 The Discrete Decomposition

In order to extend the HHD to networks we need to decide what to decompose. The natural choice is an edge flow.

Let $\mathcal{G}_{\rightleftarrows} = \{\mathcal{V}, \mathcal{E}_{\rightleftarrows}\}$ be a network consisting of vertices $\mathcal{V}$ and edges $\mathcal{E}_{\rightleftarrows}$. Assume that the network is finite and connected, and does not include multi-edges or self-loops. If the network is not connected then the same analysis can be applied to each separate component independently. Assume that all edges are directed, i.e. there is at most one edge pointing from one vertex to another. In addition, assume that, if there is an edge from i to j, then there is necessarily an edge from j back to i. Then for every pair of directed edges there is an undirected edge. Let $\mathcal{G}$ be the undirected version of $\mathcal{G}_{\rightleftarrows}$ with one undirected edge for every pair of directed edges in $\mathcal{E}_{\rightleftarrows}$ [32]. Let $V = |\mathcal{V}|$ be the number of vertices and $E = |\mathcal{E}| = |\mathcal{E}_{\rightleftarrows}|/2$ be the number of edges in the undirected version of $\mathcal{G}_{\rightleftarrows}$. Equivalently, E is the number of pairs of connected vertices.

N.b. NOT on the vertices.

Consider a function f on the directed edges of a network that maps from edges in $\mathcal{E}_{\rightleftarrows}$ to $\mathbb{R}$. This function is alternating if $f_{ij} = -f_{ji}$ where i, j are a pair of connected vertices. An alternating function on the edges is an edge flow [16].

Edge flows are analogous to vector fields in the continuum. Throughout this dissertation we will treat edge flows as vectors in $\mathbb{R}^E$. This is accomplished by introducing a reference orientation for each undirected edge. The corresponding directed graph is an orientation of $\mathcal{G}$ [33]. Index the undirected edges from 1 to E. Then for edge k let $i(k)$ and $j(k)$ represent the start of the edge and the end of the edge. Then $f_k = f_{i(k)j(k)}$, so the flow

$f \in \mathbb{R}^E$. Choosing which endpoint of an edge is the start and which is the end is an arbitrary sign convention. This sign convention defines a reference orientation for the flow. If f_k is positive then the flow is in the direction assigned by the reference orientation, and if f_k is negative then the flow points backwards against the reference orientation. Note that, if f is written with a single index then the index refers to the edge, and if f is indexed with a pair of indices then the first index is the start of the edge and the second is the end of the edge.

Next, we need a discrete equivalent to the gradient operator. The gradient should map from the space of scalar functions on the vertices to edge flows. Let G be the $E \times V$ matrix such that:

$$[Gu]_k = u_{j(k)} - u_{i(k)}. \tag{2.9}$$

N.b., so long as the number of edges differs from the number of nodes, the matrices G, D, etc. are not square.

This requires that the ki entry of G is zero unless i is the start or end of edge k. If i is the start of edge k then $G_{ki} = -1$ and if i is the end of edge k then $G_{ki} = 1$. Therefore, G is equal to the edge indicidence matrix for the directed network given by orienting each undirected edge according to the chosen reference orientation.

The divergence operator can also be easily generalized. The divergence of a vector field is the net flow out of any point in the field. So, by analogy the divergence at a node in the network should be the net flow out of that node. Therefore the divergence should map from the space of edge flows to scalar functions on nodes. Let $\mathcal{N}_i$ be the neighborhood of vertex i. Then the divergence D is the $V \times E$ matrix such that:

$$[Df]_i = \sum_{j \in \mathcal{N}_i} f_{ij}. \tag{2.10}$$

Then:

$$D = -G^\intercal. \tag{2.11}$$

The fact that the divergence is the (negative) transpose of the gradient can be checked easily. The i^{th} column of G is zero for all edges k that do not connect to node i, is one for all edges arriving at i, and negative one for all edges leaving i. Therefore $-[G^\intercal f]_i = \sum_{k|i(k)=i} f_k - \sum_{k|j(k)=i} f_k = \sum_{k|i(k)=i} f_{ij(k)} - \sum_{k|j(k)=i} f_{i(k)i}$. The first sum runs over all edges leaving i, and the second runs over all edges arriving at i. By the alternating property of the edge flow $-f_{i(k)i} = f_{ii(k)}$ so the product at edge i is the sum of f_{ij} over all $j \in \mathcal{N}_i$.

Since G and D are often large and sparse it is often easier to work with G and D implicitly rather than explicitly.

The curl is a little less obvious. The curl at a point in a continuous vector field is the tendency of the field to circulate about the point on infinitesimally small cycles. This is defined by shrinking path integrals over closed cycles about the point. Since the domain is continuous we can take a cycle to zero anywhere in the domain, so have a curl for every point in the domain. This property does not hold on a network. There is no way to shrink a cycle of states infinitesimally small since no cycle can ever include fewer than three states. It follows that the discrete curl cannot map to functions whose domain is the states in the network. Instead the curl must map to some set of functions whose domain are some set of elementary cycles.

Let $\mathcal{C}$ be a collection of simple closed loops, containing $|\mathcal{C}|$ loops. Assign each loop a reference orientation. Then the curl operator assocated with the set of loops, $C(\mathcal{C})$ is a $|\mathcal{C}| \times E$ matrix. Let $\mathcal{C}_l$ be the sequence of nodes $\{i_1, i_2, ...i_{|\mathcal{C}_l|} = i_1\}$ when precessed in its positive direction. Then:

$$[C(\mathcal{C})f]_l = \sum_{h=1}^{|\mathcal{C}_l|-1} f_{i_h i_{h+1}}. \tag{2.12}$$

Then $[C(\mathcal{C})f]_l$ is analogous to the path integral of f around the cycle $\mathcal{C}_l$.

Lim and Jiang [15, 16] choose to use the set of connected three-cliques of nodes (tri-

angles) to define the curl. This restricts the discrete curl to triangles, so misses circulation on larger loops. We do not consider triangles more fundamental than larger loops, so will define the curl using a more general set of elementary cycles. Instead we will require that $\mathcal{C}$ is a cycle basis.

A cycle basis is a collection of irreducible simple closed circuits that cannot themselves be decomposed, but can be combined to produce all other cycles in the network [33, 34, 35]. Formally, a cycle is is a subgraph of $\mathcal{G}$ for which, depending on the author, all nodes either have degree 2, or even degree [36]. A circuit is a cycle where every node in the cycle neighbors exactly two other nodes in the cycle, and the cycle is connected [33]. Linear combination of cycles is defined by the following rule: include all edges that appear in one of the cycles but not both. That is, the sum of two cycles is the symmetric difference of the edge sets of the two cycles [37]. The set of cycles of a graph form a vector field with addition defined via symmetric difference, or, integer addition modulo 2 [37][1]. A cycle basis is the basis for the cycle space of the network [34]. In electric circuit theory a cycle basis is called a basis mesh, and the study of cycle bases can be traced back to Kirchoff who introduced fundamental cycle bases when introducing his famous circuit laws [38]. Cycle bases appear across a wide variety of applications and are a classical topic in graph theory (see [33] for a helpful review of cycle bases).

Most graphs do not admit a unique cycle basis, and as a consequence there are many different classes of cycle bases each with different properties [35]. Considerable effort has been devoted to the efficient construction of cycle bases that lie in a specific classes, or that minimize some undesirable property, both for general networks, and for specific classes of networks [39, 40, 41, 42, 43, 44].

[1] In some applications cycle bases are defined with respect to a different addition operation and field (cf. [33]).

Cycle bases are most intuitive on planar graphs. The set of faces of the network, excluding one face, is a cycle basis for any planar network [37, 45]. More generally, a cycle basis is "sparse" or a "2-basis" if no edge in the network is included in more than two cycles. A network is planar if and only if it has a sparse cycle basis [35]. For this reason a cycle 2-basis is also referred to as a planar basis.

Any finite connected network admits a cycle basis. This fact can be proved constructively using a spanning tree. A spanning tree is a connected subgraph of $\mathcal{G}$ consisting of all vertices in $\mathcal{V}$, but with only a subset of the edges. The subset of edges must form a tree, $\mathcal{T}$, which spans the graph (connects all vertices) and does not contain any loops [33]. Any edge not contained in the tree is chord. Suppose edge k is a chord. Then, since $\mathcal{T}$ is a spanning tree there is a unique path from $j(k)$ to $i(k)$ in the tree. Therefore, if edge k is added back into the tree, the combined network will contain exactly one cycle. That cycle is the fundamental circuit associated with the chord [33]. Thus each chord left out by the tree is associated with a cycle in the original graph. Moreover, these cycles are necessarily independent since the cycle associated with chord k never contains any of the other chords. Given a spanning tree $\mathcal{T}$ the set of cycles formed by adding each chord to the tree is a cycle basis. Cycle bases associated with the chords of a spanning tree are fundamental cycle bases [33, 37].

The dimension of the cycle space, sometimes called the cyclomatic or Betti number [15, 35, 46], is the number of cycles needed to form a cycle basis. This dimension can be easily computed using the spanning tree construction for fundamental cycle bases. A fundamental cycle basis contains one cycle for each chord left out of the original network by the tree $\mathcal{T}$. The tree $\mathcal{T}$ connects V vertices. Any tree with V vertices has $V-1$ edges. The original network has E undirected edges, therefore the tree leaves out $E-(V-1)=E-V+1$ chords. Then $L=E-V+1$ is the dimension of the cycle space. Here we

use L for "loop" to denote the dimension of the cycle space. Note that while there is not a unique cycle basis for $L > 1$, all cycle bases will include L cycles [33].

Given a cycle basis we have the basic topology necessary to define a curl operator on undirected networks. Let $\mathcal{C}$ be a cycle basis for the graph $\mathcal{G}$, in which each pair of directed edges in $\mathcal{G}_{\rightleftarrows}$ is replaced with an undirected edge. Assign an arbitrary reference orientation to each cycle in the cycle basis. When possible, cycle orientation should be chosen so that cycles sharing an edge cross it in opposite directions. This is possible by orienting all cycles in a planar graph either clockwise or counterclockwise, but is not possible for all cycles in non-planar graphs. Then the curl is the $L \times E$ matrix defined by Equation (2.12). The lk entry of the curl is zero if edge k is not included in loop l, is one if loop l crosses edge k in its forward direction, and negative one if loop l crosses edge k in its backward direction. The curl maps from the space of edge flows to the space of alternating functions on the chosen cycle basis. If the edge flow represents the work required to cross an edge, then the curl measures the work to circumnavigate each basis cycle.

The adjoint curl is defined by taking the transpose of the curl operator. The adjoint curl maps from the space of alternating functions on cycles to the space of alternating functions on edges. In some sources the curl transpose is the cycle matrix associated with the basis $\mathcal{C}$ [33].

In order to complete the decomposition we need to show that the range of the gradient is in the nullspace of the curl (see Equation (2.3) and Equation (2.8)). This is trivial. If f is in the range of the gradient then there exists a scalar function on the nodes, u, such that $f = Gu$. But then the sum of f over any path is telescoping since $f_{ij} + f_{jk} = (u_j - u_i) + (u_k - u_j) = u_k - u_i$. Therefore the sum of f over any path is the difference in u evaluated at the endpoints of the path. If the path is a cycle then this difference is zero since the path ends where it starts. The curl evaluates the sum of f around the cycles in the

cycle basis, therefore $Cf = 0$ if $f \in \text{range}\{G\}$. Thus:

$$CG = 0, \text{and } \text{range}\{G\} \subseteq \text{null}\{C\}. \tag{2.13}$$

Transposing, the adjoint curl is contained in the nullspace of the divergence: $DC^\intercal = -G^\intercal C^\intercal = -(CG)^\intercal = 0^\intercal$.

It follows immediately from Equation (2.5) that:

Lemma 3 (The Discrete HHD). *If $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is a finite connected undirected network, then if G is the discrete gradient operator, and C is a discrete curl operator, the space of edge flows $\mathbb{R}^E$ can be decomposed:*

$$\mathbb{R}^E = \text{range}\{G\} \oplus (\text{null}\{D\} \cap \text{null}\{C\}) \oplus \text{range}\{C^\intercal\} \tag{2.14}$$

where $D = -G^\intercal$ is the divergence operator, and any edge flow $f \in \mathbb{R}^E$ can be decomposed:

$$f = -G\phi + h + C^\intercal\theta \tag{2.15}$$

where $\phi \in \mathbb{R}^V$ is a scalar-valued function on the vertices, h is a harmonic edge flow in the null space of the Hodge-Laplacian $C^\intercal C + GG^\intercal$, and $\theta \in \mathbb{R}^L$ is an alternating function on the cycles of the cycle basis.

Here ϕ is analogous to the scalar potential, and θ is analogous to the vector potential in the continuum. For a finite network with closed boundaries (no edges leaving the network) the harmonic component h is necessarily zero. This condition will guarantee a unique decomposition into just a conservative component, and a rotational component.

Lemma 4. *If $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is a finite connected undirected network, and C is defined using a fundamental cycle basis, then* range$\{G\}$ = null$\{C\}$ *so the harmonic component of the HHD defined in Equation* (2.15) *is zero.*

Proof. The orthogonality of C and G ensure range$\{G\} \subseteq$ null$\{C\}$. Therefore, to show range$\{G\}$ = null$\{C\}$ it is sufficient to show that null$\{C\} \subseteq$ range$\{G\}$. That is, to show that $Cf = 0$ implies there exists a scalar function on the nodes, $u \in \mathbb{R}^E$, such that $f = Gu$. Let $\mathcal{T}$ denote the spanning tree associated with the fundamental cycle basis. Let $u(i)$ be the function defined by summing the edge flow over the path in $\mathcal{T}$ from the first vertex in the graph to vertex i. Our goal is to show that if $Cf = 0$ then $f = Gu$. By construction $[Gu]_k = f_k$ on any edge k in the tree $\mathcal{T}$. This leaves only the chords. Consider chord k from node $i(k)$ to node $j(k)$. Since $Cf = 0$, the sum of f around the cycle in $\mathcal{C}$ containing chord k equals zero. It follows that $-f_{i(k)j(k)}$ equals the sum of f over the path from $j(k)$ to $i(k)$ in the tree. Since $f_k = [Gu]_k$ for edges in the tree, the sum of f over paths in the tree is telescoping in u. It follows that the sum over the path is $u_{i(k)} - u_{j(k)}$. But since the curl of f is zero this requires $-f_{i(k)j(k)} = -f_k = u_{i(k)} - u_{j(k)} = -[Gu]_k$ on all chords. Since $f_k = [Gu]_k$ on all edges of the tree, and all chords, $f = Gu$ on all edges in the network. Therefore, if $f \in$ null$\{C\}$ then $f \in$ range$\{G\}$ so null$\{G\}$ = range$\{C\}$.

This equality implies the space of harmonic flows, (null$\{G^\intercal\} \cap$ null$\{C\}$) only contains $h = 0$ since null$\{C\}$ = range$\{G\}$ and the range of any operator is orthogonal to the nullspace of its transpose.

□

Corollary 4.1. *If $\mathcal{G}$ is a finite connected undirected network and the curl C is defined with respect to a cycle basis such that there exists an invertible matrix T so that $C = T\hat{C}$ where*

$\hat{C}$ is the curl for a fundamental cycle basis, then if $C^\intercal f = \mathbf{0}$, then the curl of f around any cycle is zero.

Proof. First, if $C = T\hat{C}$ for an invertible matrix T then null$\{C\}$ = null$\{\hat{C}\}$. By assumption $\hat{C}$ is the curl defined with respect to a fundamental cycle basis so null$\{\hat{C}\}$ = range$\{G\}$. It follows that null$\{C\}$ = range$\{G\}$. Therefore, if $f \in$ null$\{C\}$ then $f \in$ range$\{G\}$ so there exists a potential function ϕ such that $f = -G\phi$. Then, the sum of f over any path is the difference in potential at either end, so the curl of f around any cycle is zero.

□

We are now prepared to state the key result that grounds the entire dissertation.

Theorem 5 (The Discrete HHD for Finite Networks)**.** *Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ be a finite connected undirected network. If G is the discrete gradient operator and C is a discrete curl operator defined with respect to a cycle basis such that there exists an invertible matrix T so that $C = T\hat{C}$ where $\hat{C}$ is the curl for a fundamental cycle basis, then the space of edge flows $\mathbb{R}^E$ can be decomposed:*

$$\mathbb{R}^E = \text{range}\{G\} \oplus \text{range}\{C^\intercal\} \tag{2.16}$$

and any edge flow $f \in \mathbb{R}^E$ can be decomposed:

$$f = f_{\text{con}} + f_{\text{rot}} = -G\phi + C^\intercal\theta \tag{2.17}$$

where the components $f_{\text{con}} \in$ range$\{G\}$ and $f_{\text{rot}} \in$ range$\{C^\intercal\}$ are unique, $\phi \in \mathbb{R}^V$ is unique up to the addition of a constant, and θ is unique.

Proof. First, if $C = T\hat{C}$ for an invertible matrix T then null$\{C\}$ = null$\{\hat{C}\}$. By assumption $\hat{C}$ is the curl defined with respect to a fundamental cycle basis so null$\{\hat{C}\}$ = range$\{G\}$.

It follows that null$\{C\}$ = range$\{G\}$. Then the harmonic component is necessarily zero and the subspaces range$\{G\}$ and range$\{C^\intercal\}$ are orthogonal. It follows that:

$$\mathbb{R}^E = \text{null}\{C\} \oplus \text{range}\{C^\intercal\} = \text{range}\{G\} \oplus \text{range}\{C^\intercal\}$$

and that the decomposition of f into $f_{\text{con}} \in \text{range}\{G\}$ and $f_{\text{rot}} \in \text{range}\{C^\intercal\}$ is unique.

The scalar function on the nodes, ϕ, is unique up to the addition of a constant since G has a one-dimensional nullspace associated with the vector of all ones, $\mathbf{1}$. It is trivial to see that $G\mathbf{1} = 0$ since $1 - 1 = 0$. The nullspace of G only contains vectors whose entries are all identical (vectors proportional to $\mathbf{1}$). Suppose $Gu = 0$ for some vector u. Then $[Gu]_k = u_{j(k)} - u_{i(k)} = 0$ implies $u_{j(k)} = u_{i(k)}$ for all edges k. Thus any pair of nodes that are connected by a path must have $u_i = u_j$. Since we assumed that $\mathcal{G}$ is connected $u_i = u_j$ for any i and j.

The alternating function on the cycle basis, θ, is unique if the nullspace of $C^\intercal$ contains only $\theta = \mathbf{0}$. Since T is invertible, null$\{C^\intercal\}$ = null$\{\hat{C}^\intercal T^\intercal\}$ = null$\{\hat{C}^\intercal\}$. Therefore, if null$\{\hat{C}\} = \{\mathbf{0}\}$ then θ is unique. To show null$\{\hat{C}\} = \{\mathbf{0}\}$ we show that $\hat{C} \in \mathbb{R}^{L \times E}$ has rank L. Recall that $\hat{C}$ is defined with respect to a fundamental cycle basis. Reindex the edges so that the chords defining the cycle basis are all indexed first. Each chord appears in exactly one basis cycle, therefore the first $L \times L$ block of $\hat{C}$ is diagonal with diagonal entries ± 1. Therefore, the $\hat{C}$ has rank L, the L columns of $\hat{C}^\intercal$ are linearly independent, and null$\{\hat{C}^\intercal\} = \{\mathbf{0}\}$.

□

It is worth pausing to interpret this conclusion. Theorem 5 establishes that any edge flow f on a finite network can be uniquely decomposed into two components, one which is conservative and the other which is rotational. The conservative component can be

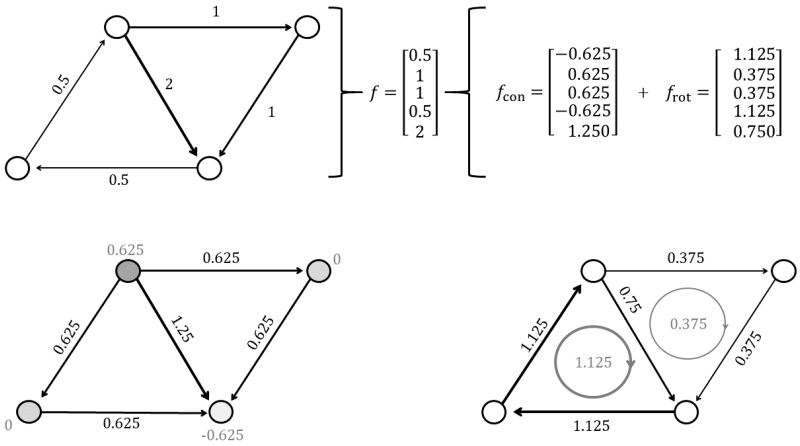

Figure 2.1: An example Helmholtz-Hodge Decomposition of an edge flow on a pair of linked triangles. The figure in the upper left shows the original edge flow. The edge flow is represented with the black numbers next to each edge. The corresponding vector f, and its conservative and rotational components are shown along the top. The figure in the bottom left shows the conservative component of the flow. The values of the scalar potential, ϕ, at each node is shown in grey. The figure in the bottom right shows the rotational component of the flow, and the rotational potential ϕ assigned to each loop is shown in grey.

expressed as the negative gradient of a scalar function, ϕ, defined on the nodes of the network. This is analogous to a scalar potential in the continuum (see Equation (2.7)). Like the scalar potential, it is only uniquely determined up to the addition of a constant since it is the difference in potential, not absolute potential, which determines the flow. The rotational component can be expressed as the curl transpose applied to an alternating function, θ, defined on the loops in the chosen cycle basis. This is analogous to the vector potential, though we will avoid using the name vector potential since θ, like ϕ, is scalar valued. Instead we will refer to θ as the rotational potential since it drives rotation. An example decomposition is shown in Figure 2.1.

Note that we defined the discrete gradient in terms of how it mapped a scalar function on the nodes to an edge flow (see Equation (2.9)). In contrast, the curl was defined as a mapping from edge flows, to the space of loops. Since the rotational field is defined by

$C^\intercal\theta$ it is essential to understand how the transpose maps from an alternating function on the loops to an edge flow.

The product $C^\intercal\theta$ is a linear combination of the columns of $C^\intercal$. Each column of $C^\intercal$ is an edge flow around a basis loop. The l^{th} column of $C^\intercal$ corresponds to a unit flow around the l^{th} basis loop in its forward direction. Therefore, $C^\intercal\theta$ is a linear combination of unit edge flows around each loop in the cycle basis. The value of the rotational flow on edge k, $[C^\intercal\theta]_k$, is the sum of θ_l over all loops l including edge k in its positive direction, minus the sum of θ_l over all loops l including edge k in its negative direction. Note that, since $CG = 0$ the divergence of the adjoint curl, $DC^\intercal = -(CG)^\intercal$, equals 0. It follows that the divergence of the rotational flow, $f_{\text{rot}} = C^\intercal\theta$, is always zero.

2.2.3 Subspaces and Operators

Elementary Properties of the Subspaces

What is the dimension of the space of conservative edge flows, range$\{G\}$, and the space of rotational edge flows, range$\{C^\intercal\}$?

In the proof of Theorem 5, we showed that the gradient has a one-dimensional nullspace corresponding to vectors whose entries are all the same. Since $G \in \mathbb{R}^{E\times V}$ it follows that range$\{G\}$ has $V - 1$ dimensions. In the proof of Theorem 5, we also showed that the curl has rank L where $L = E-(V-1)$ is the dimension of the cycle space. Since $C^\intercal \in \mathbb{R}^{E\times L}$ it follows that range$\{C^\intercal\}$ has $L = E - (V - 1)$ dimensions. Thus the decomposition into the conservative and rotational subspaces decomposes an edge flow with E degrees of freedom into an edge flow with $V - 1$ degrees of freedom, and an edge flow with $L = E - (V - 1)$ degrees of freedom.

Since $CG = 0$ the conservative subspace is orthogonal to the rotational subspace, all

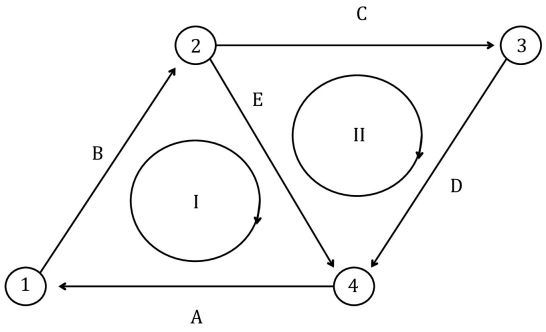

Figure 2.2: Pair of triangles sharing an edge.

conservative flows are curl-free, and all rotational flows are divergence free. Therefore the HHD is a decomposition onto two orthogonal subspaces, the conservative flow is the projection of the original flow onto the conservative subspace, and the rotational flow is the projection of the original flow onto the rotational subspace.

Change of Cycle Basis

Decomposing the network into a conservative and rotational part required defining a domain for both the scalar potential ϕ and the rotational potential θ. The domain of ϕ is the set of vertices, so is unique and fixed by $\mathcal{V}$. The domain of θ is a particular cycle basis for the network. All networks admit a cycle basis, however the choice of cycle basis is not, in general, unique. This leads to the natural question, how does the rotational potential θ depend on the choice of basis? And, implicitly, does the choice of cycle basis effect the scalar potential ϕ?

Consider a pair of linked triangles sharing a common edge as shown in Figure 2.2. The natural cycle basis for this pair is shown in the figure. In this basis the first loop moves clockwise from node 1 to node 2 to node 4, and the second loop moves clockwise from node 2 to node 3 to node 4.

Alternatively we could define a cycle basis using the loop from 1 to 2 to 4, and from 1

to 4 to 3 to 2. That is, if I denotes the first loop, and II denotes the second loop, we can define a cycle basis $\{I, II\}$ or a cycle basis $\{I, -(I + II)\} = \{I, III\}$.

By definition, any loop in a network can be expressed as a linear combination of a set of basis cycles. If a network admits multiple cycle bases, any cycle basis can be expressed as a linear transformation of any other cycle basis.

In the example discussed above the linear transform is:

$$T = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}.$$

Here T is invertible with dimension equal to the dimension of the cycle space.

Now suppose we are given two cycle bases whose curl is related by a transform T by:

$$\hat{C} = TC. \tag{2.18}$$

If the network is planar and $\mathcal{C}$ is a planar cycle basis, then the linear transform T will match the linear combination of cycles used to define $\hat{\mathcal{C}}$ given $\mathcal{C}$. This equivalence can be verified for the example. The first cycle basis has curl:

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

and the second basis has curl:

$$\hat{C} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 & 0 \end{bmatrix}.$$

Then it is easy to verify that:

$$\hat{C} = TC.$$

The matrix T mapping from $\hat{C}$ to C may not match the linear transformation used to move between $\hat{\mathcal{C}}$ and $\mathcal{C}$ for all possible pairs of cycle bases since the rule for linear combination of cycles is based on the symmetric difference of pairs of loops. This addition rule is equivalent to the addition of integers modulo 2. In contrast the transform TC uses the standard definition of addition. Provided only two cycles are combined then the sum of two cycles in a cycle basis, $\mathcal{C}_l + \mathcal{C}_h$, amounts to adding or subtracting one row of the curl from another. If the two cycles have the same orientation on their shared boundary then one row should be subtracted from the other. If the two cycles have different orientations on their shared boundary then the two rows of the curl should be added. Therefore, if $\hat{\mathcal{C}}$ can be reached from $\mathcal{C}$ by a sequence of combinations of pairs of cycles then C' can be reached from C by a sequence of products with a series of matrices, each corresponding to the appropriate elementary row operation. The product of these row operations, T, will match the transform from $\mathcal{C}$ to $\hat{\mathcal{C}}$. If $\hat{\mathcal{C}}$ cannot be reached by a sequence of pairwise combinations of loops then T may not match the linear transform between the cycle spaces.

Regardless the choice of basis, Theorem 5 guarantees that the range of the curl is the nullspace of the divergence. The nullspace of the divergence is the nullspace of G^{T}, which is uniquely defined and does not depend on the choice of cycle basis. It follows that the range of the curl is independent of the choice of cycle basis. Therefore, if C and $\hat{C}$ are the curl operators associated with two different cycle bases they are both matrices of the same size that have the same range, so there must exist an invertible linear transform T such that $TC = C'$.

Since the range of the gradient and the curl are both independent of the choice of cycle

basis, the projection of the edge flow f onto either subspace is independent of the choice of cycle basis. Therefore f_{con} and f_{rot} are independent of the choice of cycle basis. Since $-G\phi = f_{\text{rot}}$, the scalar potential ϕ is also independent of the choice of cycle basis. This leaves only rotational potential θ. Since f_{rot} is independent of the choice of cycle basis:

$$C^{\intercal}\theta = f_{\text{rot}} = \hat{C}^{\intercal}\hat{\theta} = C^{\intercal}T^{\intercal}\hat{\theta}.$$

We have shown that $C^{\intercal}$ is always full rank so $C^{\intercal}\theta = C^{\intercal}T^{\intercal}\hat{\theta}$ implies:

$$\theta = T^{\intercal}\hat{\theta}, \quad \hat{\theta} = T^{-\intercal}\theta. \tag{2.19}$$

Combined, these two results give a simple rule for moving between cycle bases. Given two cycle bases whose curl is related by a linear transform T, the only component of the HHD which changes is θ, which is replaced with $\hat{\theta} = T^{-\intercal}\theta$.

Operator Duality for Planar Graphs

Here we illustrate that the curl and gradient operators are dual on planar graphs. Dual relations of this kind have been used by some authors to investigate cyclic fluxes at the steady state of Markov chains [46].

Suppose that $\mathcal{G}$ is a planar graph. Then the set of faces of $\mathcal{G}$, excluding any single face, is a cycle basis. Note that the set of faces of $\mathcal{G}$ includes the exterior when $\mathcal{G}$ is embedded in the plane. A basis of this form is an example of a sparse cycle basis or 2-basis: a cycle basis such that no edge in the network is included in more than two basis cycles [35]. The reference orientation for each face on the graph can be chosen so that any edge shared by two loops is crossed in opposite directions by those loops. This fact can be proved by noting that, if two loops share an edge and both cross the edge in the same direction then

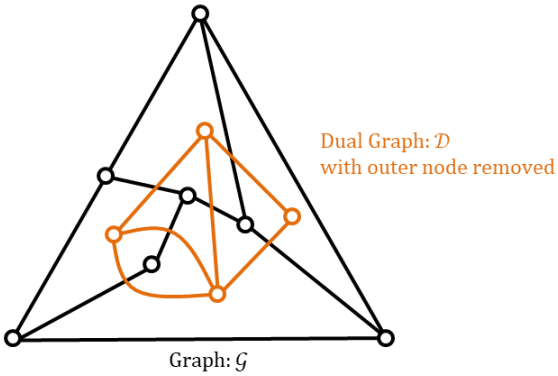

Figure 2.3: A planar graph $\mathcal{G}$ and the corresponding dual $\mathcal{D}$ with the node corresponding to the exterior face removed.

reversing the orientation of one loop, and removing any shared edges, leaves another larger cycle all traversed in the same direction. Since we assumed the graph is planar none of the other faces are changed if the conflicting edge is removed. Therefore, we can start by arbitrarily orienting a particular loop, then choosing the orientation on a neighboring face so that the two loops cross their shared boundary in opposite directions. Remove the shared boundary from the network, and merge the two loops into one. Now iterate the process, merging loops until every face in the graph is assigned a consistent reference orientation.

The dual graph associated with a planar graph has a node for every face in the original planar graph and an edge for every edge in the original graph. The edges in the dual connect the nodes corresponding to faces on opposite sides of the corresponding edges of the original graph. Thus, if an edge in the original graph separates two faces h and l, there is a matching edge in the dual connecting nodes h and l. Note that if the faces h and l share multiple edges there may be multiple edges connecting nodes h and l in the dual. Also note that if a face of the original graph appears on both sides of the same edge then the dual will contain a self-loop connecting the corresponding node to itself. An example planar graph, and its dual with the node corresponding to the exterior face removed is illustrated

in Figure 2.3.

Let $\mathcal{D}$ denote the dual of $\mathcal{G}$. The gradient of the dual, $G_{\mathcal{D}}$, is $E \times (L+1)$, and where the k^{th} row of $G_{\mathcal{D}}$ is equal to $e_{h(k)} - e_{l(k)}$ if edge k of the original graph separates loops h and l. Note that, some of the rows of G may be identical since the dual may contain multiple edges between the same pairs of nodes, and some may be empty since the dual may include self-loops. Note that if an edge is bordered on opposite sides by the same loop then the corresponding edge in the dual is a self-loop, so the corresponding row of the gradient is all zeroes.

Let $\hat{C}$ denote the curl of the original graph with one row added corresponding to the face removed from the cycle basis. Then $\hat{C}^\intercal$ is the gradient operator for the dual.

First, $C^\intercal$ is $E \times L$ so $\hat{C}^\intercal$ is $E \times (L+1)$. Next, the k^{th} row of $\hat{C}^\intercal$ contains only zeros, ones, and negative ones. Suppose edge k separates loops h and l and $h \neq k$. Since no edge borders more than two basis cycles, and all basis cycles that share an edge cross it in opposing directions, $[\hat{C}^\intercal \hat{\theta}]_k$ is either $\theta_h - \theta_l$ or $\theta_l - \theta_h$ depending on the choice of reference orientation. If an edge borders the same loop twice then the corresponding entry of $\hat{C}^T\theta = 0$. Then the k^{th} row of $\hat{C}^\intercal$ will be all zeros if edge k is bordered on both sides by the external face. This matches the gradient of the dual, whose k^{th} row is all zeros if the corresponding edge in the dual is a self-loop. Therefore $\hat{C}^\intercal$ is the same as the gradient of the dual.

Since all but one of the vertices of the dual correspond to the cycles in the cycle basis, the action of the curl transpose on θ is equal to the action of $G_{\mathcal{D}}$ on u where $u_l = \theta_l$ on all nodes l corresponding to loops in the cycle basis, and $u_{L+1} = 0$ on the node $L+1$ corresponding to the external face. Therefore the action of $C^\intercal$ on θ is the same as the action of the gradient of the dual on an equivalent scalar potential, with the convention that the scalar potential on the external face is always equal to zero.

Name	Continuous	Discrete	Mapping	Meaning	Properties
Gradient	∇	G	vertices to edges	Slope	Curl-free
Divergence	$\nabla\cdot$	D	edges to vertices	Flux out	$D = -G^{\mathsf{T}}$
Curl	$\nabla\times$	C	edges to loops	Cycling flux	...
Adjoint Curl	$\nabla^*\times$	C^{T}	loops to edges	Rotation	Divergence-free

Table 2.2: The operators.

Divergence Rule, Product Rules, and Stoke's Theorem

Here we show that the network operators retain some of the important properties of the analogous differential operators. We have already shown they are mutually orthogonal, so admit a HHD. Since the operators are all linear they automatically inherit the linearity of ∇. This section will show they also obey: the divergence theorem, Stokes' theorem, and analogous product rules. These relations lay the groundwork for solving Poisson's equations using Green's functions. They also strengthen the analogy between the two classes of operators, and help carry intuition from continuous domains to the network potentials. Note that the results for the curl are restricted to planar graphs since they depend on a geometric notion of interior that is not defined for general graphs. The results for the gradient and divergence apply to all networks. A review of the operators is provided in Table 2.2.

First, the discrete divergence, D, obeys the divergence theorem:

Lemma 6 (Divergence Theorem)**.** *Given a set of nodes Ω, denote the set of edges that pass from nodes inside the set to nodes outside $\partial\Omega$. Let n_k equal 1 if edge k leaves Ω and -1 if edge k enters Ω. Then the total flux out of Ω is the sum of the divergence of f over all nodes in Ω:*

$$\sum_{i\in\Omega}[Df]_i = \sum_{k\in\partial\Omega} f_k n_k. \tag{2.20}$$

Proof. The divergence at a node is the sum of f pointing out from the node (net flow out of the node). Suppose we sum the divergence of two neighboring nodes i and j. Without loss of generality assume that f_{ij} is positive, and denotes a flow from i to j. Since f_{ij} leaves the first node it is added to the divergence of the first node. Since f_{ij} arrives at the second node it is subtracted from the divergence of the second node. Any flow between two nodes in the domain is either added to the first and subtracted from the second, or visa versa. Therefore all edges inside the domain Ω do not contribute to the total divergence. Since all internal flows cancel, only the flow passing through the boundary of Ω remains. Therefore the net divergence of a set equals the flux through the boundary of the set. □

For planar graphs the discrete curl, C, obeys Stokes' theorem if the cycle basis is chosen to match the faces of the graph:

Lemma 7 (Stokes' Theorem)**.** *Suppose the network is planar. Given a set of basis cycles Ω, define the interior of the set to be all nodes in the set which exclusively neighbor other nodes in Ω, and the boundary of Ω to be all nodes in Ω but not in its interior. Let $\mathcal{C}$ denote the set of all cycles contained in Ω. Let C be defined using the planar cycle basis (faces of the planar graph excluding the exterior face), with reference orientations chosen so that all edges neighboring two cycles are traversed in opposite directions by those cycles. Let $\mathcal{B}$ be the set of edges that connect nodes in the boundary set, and let n_k be 1 if the cycle contained inside Ω neighboring edge k crosses edge k in its forward direction, and -1 if it crosses edge k in its backwards direction. Then:*

$$\sum_{\mathcal{C}_l \in \mathcal{C}} [Cf]_l = \sum_{k \in \mathcal{B}} f_k n_k. \tag{2.21}$$

Proof. On a planar graph it is always possible to pick a sign convention such that all

neighboring cycles traverse their shared boundaries in opposite directions. If the graph is embedded in the plane then if all cycles are traversed clockwise or all cycles are traversed counterclockwise, then all edges neighboring two cycles are traversed in opposite directions by those cycles.

The curl over a cycle is the total work to move around the cycle in a specified direction. This is computed by summing f around the cycle. Two neighboring cycles are necessarily separated by some path. We chose the sign convention so that two neighboring cycles will traverse their shared boundary in opposite directions. It follows that the sum of the curls of neighboring loops cancels any contribution from their shared boundary. Thus the sum over a set of neighboring loops leaves only the contribution to the curl from the edges on the boundary $\mathcal{B}$.

□

Lemma 8 (Gradient Product Rule)**.** *Given two functions on the nodes ψ, ϕ the gradient of the product $u_i = \phi_i \psi_i$ obeys:*

$$[Gu]_k = \frac{\phi_{i(k)} + \phi_{j(k)}}{2}[G\psi]_k + \frac{\psi_{i(k)} + \psi_{j(k)}}{2}[G\phi]_k. \tag{2.22}$$

Proof. The proof is entirely arithmetic. First:

$$[Gu]_k = \psi_{j(k)}\phi_{j(k)} - \psi_{i(k)}\phi_{i(k)}.$$

Then notice that:

$$\psi_j(\phi_j - \phi_i) + \phi_i(\psi_j - \psi_i) = \psi_j\phi_j - \psi_j\phi_i + \phi_i\psi_j - \phi_i\psi_i = \psi_j\phi_j - \psi_i\phi_i$$
$$\psi_i(\phi_j - \phi_i) + \phi_j(\psi_j - \psi_i) = \psi_i\phi_j - \psi_i\phi_i + \phi_j\psi_j - \phi_j\psi_i = \psi_j\phi_j - \psi_i\phi_i.$$

Combining the last two equations yields:

$$\frac{\phi_{i(k)} + \phi_{j(k)}}{2}[G\psi]_k + \frac{\psi_{i(k)} + \psi_{j(k)}}{2}[G\phi]_k = \psi_j\phi_j - \psi_i\phi_i = [Gu]_k.$$

□

Lemma 9 (Adjoint Curl Product Rule)**.** *Given a planar network, a curl operator defined using the planar cycle basis with reference orientations chosen so that any edge shared by two cycles is crossed in opposite directions, and two functions on the cycles ψ, θ the adjoint curl of the product $v_l = \psi_l\theta_l$ obeys:*

$$[C^\intercal v]_k = \frac{\theta_l + \theta_h}{2}[C^\intercal\psi]_k + \frac{\psi_l + \psi_h}{2}[C^\intercal\phi]_k. \tag{2.23}$$

Where k indexes an edge separating the set of cycles $\mathcal{C}_l$ and $\mathcal{C}_h$.

Proof. By construction, the adjoint curl of the function v defined on the cycles takes the difference of v between neighboring cycles. The dual graph to a planar graph has a node for every face of the graph, and an edge for every pair of neighboring faces. Then, since we chose the cycle basis to match the faces of the planar graph, the dual graph includes a node for every cycle in the cycle basis. Since the adjoint curl evaluates the difference between functions defined on the cycles of the cycle basis, the adjoint curl is the gradient of the dual graph. Since the discrete gradient always obeys the product rule, the adjoint curl of a planar graph must also follow the same product rule on its dual. □

Taken together, these results show that the network operators retain many of the important algebraic properties of the continuous operators.

2.3 Solution Methods and Alternative Characterizations

2.3.1 Least Squares and the Discrete Poisson Equations

If the network is finite then the harmonic component h is zero. In that case the range of the gradient and the range of the adjoint curl span $\mathbb{R}^E$, so there exists a unique (up to addition of a constant) pair of potentials ϕ and θ such that $-G\phi + C^\intercal\theta = f$. The conservative and rotational components of the flow are uniquely defined and equal the projection of the original flow onto the range of the gradient, and range of the curl transpose. Since the range of the gradient and range of the curl transpose are orthogonal this linear system is equivalent to a linear least squares problem in each potential:

$$\begin{aligned} \phi &= \text{argmin}_{u\in\mathbb{R}^V}\left\{||f + Gu||^2\right\} \\ \theta &= \text{argmin}_{v\in\mathbb{R}^L}\left\{||f - C^\intercal v||^2\right\} \end{aligned} \tag{2.24}$$

This pair of least squares problems are solved by any ϕ and θ which satisfy the associated normal equations:

$$\begin{aligned} G^\intercal G\phi &= -G^\intercal f \\ CC^\intercal\theta &= Cf. \end{aligned} \tag{2.25}$$

The same result could be reached by multiplying $-G\phi + C^\intercal\theta = f$ on either side by the divergence, $-G^\intercal$, or the curl, C. Since the ranges of the two operators are orthogonal $G^\intercal C^\intercal\theta = 0$ and $CG\phi = 0$, leaving $G^\intercal G\phi = -G^\intercal f$ and $CC^\intercal\theta = Cf$. The operator $G^\intercal G$ is the node Laplacian. The operator $CC^\intercal$ is the face Laplacian. The node Laplacian has a column and row for each node in the network, and the face Laplacian has a column and row for each cycle in the chosen cycle basis. Let $L_\mathcal{V}^2 = G^\intercal G$ and $L_\mathcal{C}^2 = C^\intercal C$. In this notation,

the potentials are solutions to a pair of discrete Poisson equations:

$$\begin{aligned} L_{\mathcal{V}}^2 \phi &= Df \\ L_{\mathcal{C}}^2 \theta &= Cf. \end{aligned} \tag{2.26}$$

By construction, the Laplacians are square, symmetric, and positive semi-definite. Each Laplacian has the same nullspace as the operator used to define it. Therefore the the node Laplacian has a one dimensional nullspace corresponding to vectors with constant entries. In contrast the face Laplacian is full rank and invertible. To solve for a unique ϕ, append an additional condition to the Laplace operator that fixes the potential at one point. Usually the potential at a chosen node is set to zero. Then the row and column of the node Laplacian corresponding to the fixed node can be removed from the Poisson equation. Removing a row and column from the node Laplacian produces an invertible $(V-1) \times (V-1)$ matrix.

Let $\hat{L}_{\mathcal{V}}^2$ denote the truncated node Laplacian. Similarly, let $\hat{D}$ denote the divergence with the row corresponding to the fixed node removed, and $\hat{\phi}$ denote the potential with the entry corresponding to the fixed node removed. Reindex the nodes so that the fixed node is indexed first. Then ϕ is uniquely specified by the discrete Poisson equation:

$$\hat{L}_{\mathcal{V}}^2 \hat{\phi} = \hat{D} f, \quad \phi = [0; \hat{\phi}]. \tag{2.27}$$

Then any constant can be added to the potential ϕ. For example, if we want $\phi_j \geq 0$ we simply subtract $\min_j \{\phi_j\}$ from ϕ. Alternatively, if we desire $\sum_j \phi_j = 0$ then we subtract the average value of the potential from ϕ.

Equation (2.25) and Equation (2.26) offer two different perspectives on the decomposition. Framing the HHD as a pair of least squares problems is useful in contexts where it is expected or desired that the edge flow will be either close to conservative, or

close to rotational. Then the conservative and cyclic components can be seen as the best approximation to the original edge flow on each subspace. Here "best" is evaluated in the l_2 sense and without weighting. In the least-squares sense f_{con} and f_{rot} are simultaneously the best conservative and best cyclic approximations to the original edge flow. This interpretation will be useful when studying the ranking of competitors in a competitive system in Chapter 4 and Chapter 5.

Framing the HHD with the discrete Poisson equations is useful since the potentials of a continuous vector field are also solution to Poisson's equations [8]. Denote the continuous Laplacian ∇^2. The continuous potentials for an arbitrary C^1 vector field $v(x)$ satisfy:

$$\begin{aligned} \nabla^2 \phi(x) &= -\nabla \cdot v(x) \\ \nabla^2 A(x) &= -\nabla \times v(x). \end{aligned} \tag{2.28}$$

The continuous Laplacian is the same for both potentials since the divergence of the gradient is, by definition, ∇^2, and the curl of the curl of the vector potential $A(x)$ is $\nabla(\nabla \cdot A(x)) - \nabla^2 A(x)$ but $A(x)$ can always be chosen so that it is divergence free.[2] In both cases $\nabla^2 = \partial_x^2 + \partial_y^2 + \partial_z^2$, and when applied to $A(x)$ the Laplacian is applied entrywise. Equation (2.28) can be reached by evaluating the divergence, or curl, of Equation (2.7).

To clarify the analogy between the discrete Laplacians and ∇^2 it is useful to consider the discrete Laplacians more closely. Consider the node Laplacian, $L_{\mathcal{V}}^2 = G^\intercal G$, first. The node Laplacian is a fundamental operator in graph theory. The spectrum of $L_{\mathcal{V}}^2$ is the heart of spectral graph theory. For example, the pseudo-determinant of $L_{\mathcal{V}}^2$ determines the number of spanning trees of the network, the dimension of the nullspace of $L_{\mathcal{V}}^2$ is the number of

[2]The convention $\nabla \cdot A(x) = 0$ is the Coulomb gauge in electromagnetism. If $A(x)$ was not divergence free, then, by the Helmholtz decomposition (Equation (2.6)), it would be possible to write $A(x)$ as the combination of a conservative vector field and a rotational vector field. But the curl of a conservative vector field is zero, so by convention we can assume that $A(x)$ does not have a conservative component. In that case $\nabla \cdot A(x) = 0$.

connected components of the graph, and the eigenvectors of $L^2_{\mathcal{V}}$ can be used for optimal embedding [47, 48, 49, 50].

The node Laplacian is defined by the product $G^{\intercal}G$. The gradient maps from functions on nodes to functions on edges. The divergence maps back from functions on edges to functions on nodes. Therefore the node Laplacian $L^2_{\mathcal{V}}$ maps from functions on nodes to functions on nodes. Suppose we define a function on the nodes. The gradient evaluates the difference in the function evaluated at either end of each edge. Now suppose we have defined a function on the edges. The divergence at a node sums over the function evaluated at all edges connected to the node. Therefore the divergence of the gradient evaluates the difference between the function evaluated at a node, and the function evaluated at its neighbors. Let d be a vector where d_i is the degree of node i and let A be the adjacency matrix of $\mathcal{G}$. Then:

$$L^2_{\mathcal{V}} = \text{diag}(d) - A \tag{2.29}$$

and:

$$[L^2_{\mathcal{V}}\phi]_i = d_i\phi_i - \sum_{j \in \mathcal{N}_j} \phi_j = d_i \left(\phi_i - \bar{\phi}_{\mathcal{N}_i}\right) \tag{2.30}$$

where $\bar{\phi}_{\mathcal{N}_i}$ is the average value of ϕ over the neighbors of node i.

Written in this manner the node Laplacian is clearly a second difference operator. This offers insight into Poisson's equation. The Laplacian of the potential is simply the difference between the potential at each node, and the average potential at its neighbors. If the potential at a node is higher than its neighbors then the Laplacian of the potential is positive. The negative Laplacian is negative, so the divergence of the associated f is positive. That is, if the potential at a node is higher then its neighbors, then f diverges away from that node. On the other hand, if the potential at a node is lower than its neighbors f converges towards that node.

Suppose the network is planar. Then leveraging the duality of the curl transpose and gradient extends the same result directly to the face Laplacian. Working through the same steps, the adjoint curl evaluates the difference in rotational potential between two neighboring faces. The curl evaluates the total flow around a loop. Together, the curl of the adjoint curl, evaluates the difference between the rotational potential on a face, and the average rotational potential on its neighbors. Again this implicit representation of $L_{\mathcal{C}}^2$ offers insight into Poisson's equation. If the rotational potential is higher at a loop than its neighbors, then f will tend to flow along the positive direction of traversal around the loop. If the rotational potential at a loop is lower then its neighbors then f will tend to flow against its positive direction of traversal. Note this interpretation only holds if the cycle basis is set to the faces of the graph, and the reference orientations are chosen consistently.

In both cases the realized flow of f, as measured by the divergence and curl, is proportional to the difference between a potential at a point and its neighbors.

The Laplacian operators can be constructed explicitly from these implicit definitions, however it is generally more efficient to work with the implicitly defined operators. Since the Laplacian operators are sparse, symmetric, and semi-positive definite they are well suited to iterative methods. The node Laplacian of simple networks are often familiar. For example the node Laplacian of a line of five nodes, and a loop of five nodes are:

$$L_{\mathcal{V}}^2 = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}, \quad L_{\mathcal{V}}^2 = \begin{bmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{bmatrix}.$$

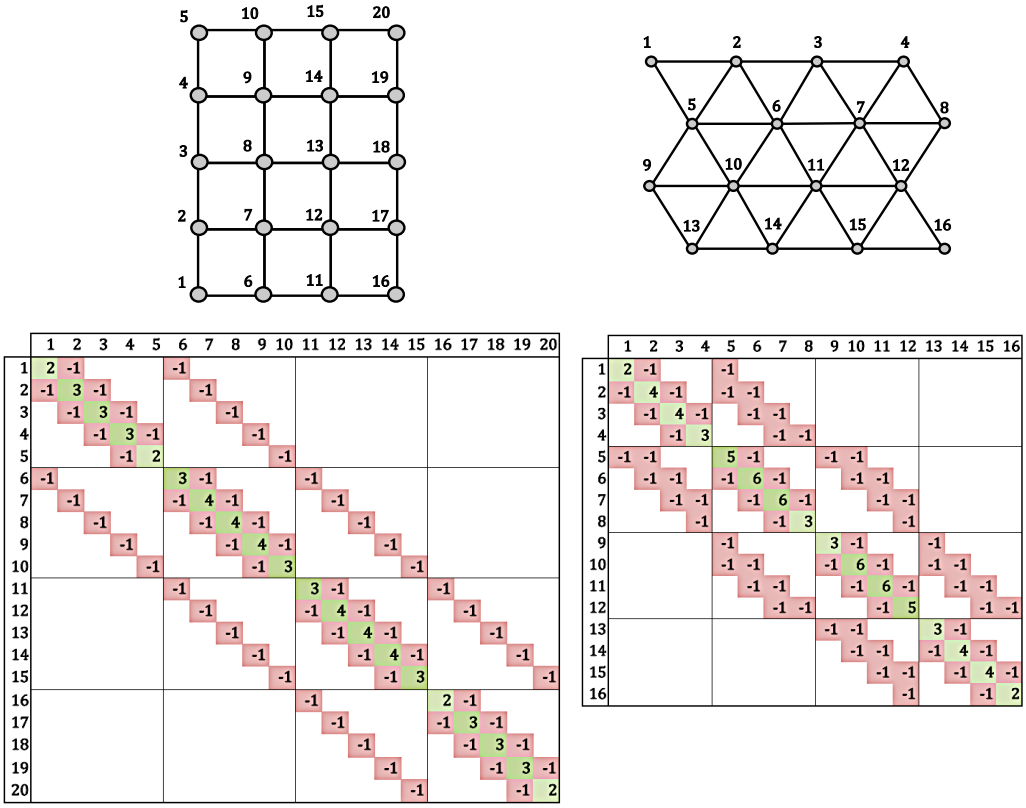

Figure 2.4: The node Laplacians for a rectangular and triangular grid. Both are symmetric, and inherit the adjacency structure of the associated network. Notice that every row and column of each Laplacian sums to zero.

Laplacians for a 5 by 4 grid, and a 4 by 4 triangular grid are shown in Figure 2.4.

The continuous Poisson equation is generally solved using Green's functions. The corresponding Green's functions are defined implicitly by $\nabla^2 G(x, x') = \delta(x - x')$ where $\delta(x - x')$ is the Dirac delta function. On a discrete domain the Dirac delta is simply a canonical basis vector $e_i = [0, 0...0, 1, 0...0]^\intercal$. It follows that the i^{th} Green's "function" is a vector $g(i)$ which satisfies $L^2 g(i) = e_i$. Assume that we have adjusted the Laplacian so that it is invertible. Then: $g(i) = (L^2)^{-1} e_i$, is the i^{th} column of the inverse Laplacian $(L^2)^{-1}$. To recover the potential sum over the right hand side of the discrete Poisson equation. Thus,

solving for the potentials via Green's functions is equivalent to solving for the columns of the inverse Laplacian, then summing over the columns. This is the direct solution to the normal equations; compute the inverse Laplacian, then multiply by the inverse to recover the potentials.

In general, forming the normal equations explicitly, and solving the associated linear system directly is neither the most efficient, nor the most stable method for solving least squares problems. Multiplying by the adjoint to form the normal equations squares the conditioning of the problem. For ill conditioned problems this can seriously reduce the accuracy of the computed solution. Worse, computing the inverse explicitly is generally very expensive. Each column of the inverse requires solving a linear system. If there are V nodes in the network then solving for each column (each Green's function) requires $\mathcal{O}\left(V^3\right)$ operations. Since the inverse Laplace operator has V nodes the method requires $\mathcal{O}\left(V^4\right)$ operations if run for each column individually. If run for the columns together, the algorithm requires $\mathcal{O}\left(V^3\right)$ operations [51]. In special cases the Green's functions may be known ahead of time, however for most networks this method is unnecessarily, or even prohibitively, expensive.

An alternative option is to solve for the conservative and rotational components first, and then solve for the potentials. The components are orthogonal projections of the original edge flow, f, onto the appropriate subspaces. The projector onto either subspace can be formed by decomposing either the gradient or the curl transpose. Since the gradient (much) is easier to build (see Section 3.5), we focus on the gradient. First, perform a QR decomposition or a singular-value decomposition of the gradient. Then identify the columns of Q or U that form a basis for the range of the gradient (columns corresponding to non-zero singular values). Then the orthogonal projector onto the range of the gradient is formed by $P_{\text{con}} = \hat{Q}\hat{Q}^\intercal = \hat{U}\hat{U}^\intercal$ where $\hat{Q}$ and $\hat{U}$ are the columns of Q and U that span

the conservative subspace. For an $m \times n$ matrix the QR decomposition requires at most $\mathcal{O}(2mn^2)$ operations. Since we only need to decompose one of the two operators the cost for computing $\hat{Q}$ is $\mathcal{O}(2EV^2)$. Using Householder triangularization instead of Graham-Schmidt reduces the cost to $\mathcal{O}(V^3)$ and is more stable [51].

Once P_{con} has been computed, the components of the HHD are given by:

$$\begin{aligned} f_{\text{con}} &= P_{\text{con}} f \\ f_{\text{rot}} &= f - f_{\text{con}}. \end{aligned} \tag{2.31}$$

Given the conservative and rotational components, the potentials satisfy $-G\phi = f_{\text{con}}$ and $C^\intercal \theta = f_{\text{rot}}$. Solving for ϕ is straightforward, and can be done directly. By definition f_{con} is conservative, so the sum of f over any path is path independent and equals the difference in the potential at the endpoints. Therefore the potential can be recovered by picking a spanning tree $\mathcal{T}$ and initial node. Set the potential at the initial node to zero, and then set ϕ_i to the sum of f over the path in $\mathcal{T}$ to node i. This only requires adding each edge in the tree once, so we only need $V - 1$ addition operations. The spanning tree can be built at the same time as the sum is performed by using a breadth first search, where the potential at each new leaf added to the tree is updated as the search progresses. This method is much more stable than solving for the potential directly using the normal equations. Path integration is equivalent to back substitution of a triangular linear system, so is backward stable [51]. The projectors have conditioning one, so moving from f to f_{con} preserves the conditioning of the original problem.

If the network is planar and the planar cycle basis is used then the rotational potential θ can be computed directly from f_{rot} by summing over a spanning tree of the dual graph. Otherwise the system $C^\intercal \theta = f_{\text{rot}}$ needs to be solved. Note that f_{rot} is, by definition, in the range of $C^\intercal$ and $C^\intercal$ has a trivial nullspace, so the linear system $C^\intercal \theta = f_{\text{rot}}$ always admits a

unique solution.

When a fundamental cycle basis is used to define the curl then $C^\intercal\theta = f_{\text{rot}}$ can be solved directly. If a fundamental cycle basis is used then each chord appears in only one loop. Therefore, if the k^{th} edge of the network is chord h, corresponding to basis cycle h, then $\theta_h = f_{\text{rot}\,k}$. Thus, when a fundamental cycle basis is used the only expensive part of performing the HHD via projection is to form the projector onto either the conservative or rotational subspaces. Note that, if the cycle basis is formed by first finding a fundamental cycle basis, then performing row operations to find a basis that is easier to interpret, the row operations can be stored and collated into a cycle basis transform T and Equation (2.19) can be used to recover the rotational potential on the desired basis from the potential on the fundamental basis.

Suppose now that the network is very large. Then it may not be possible to solve for the potentials using Green's functions or by projecting simply because it is too expensive to form the necessary operators explicitly.

That said, in most cases both the gradient and the adjoint curl are extremely sparse. The gradient only has two nonzero entries per row, and has E rows and V columns. If a network is planar then each edge connects two nodes, and separates at most two faces. Therefore, if the network has E undirected edges, the gradient and adjoint curl have $\mathcal{O}(2E)$ nonzero entries, but $\mathcal{O}(VE)$ and $\mathcal{O}(LE)$ entries in total. Therefore both the curl and the gradient become sparser the more edges, loops, or vertices in the network. Even if the graph is not planar then there always exists a cycle basis such that the curl has $\mathcal{O}(E\log(V)/\log(E/V))$ nonzero entries [33], so becomes sparse as V grows.

When the operators are sparse it may be dramatically more efficient to solve for the potentials using an iterative linear system solver. There are a variety of iterative methods for solving large, sparse linear least squares problems. These include conjugate gradient

methods (CG), biorthogonalization methods, and generalized minimum residuals (GMRES). Documentation on iterative methods is available from a variety of sources [51].

Generally these methods proceed by searching a sequence of nested subspaces. At each step they return an approximate solution. The iteration updates by choosing a search direction, and finding an optimal step along that direction. This update is usually computed via a multiplication by the linear system, so is efficient if the linear system is sparse. Since sparse matrices can often be represented implicitly, iterative methods may not require an explicit linear system, only a rule for how to apply the system. The gradient and adjoint curl are easy to define implicitly, so are well suited to iterative solvers. The associated node Laplacian also admits a simple implicit definition (see Equation (2.30)). Thus iterative methods are well suited to finding the potentials of large networks.

In general we follow [16] and use iterative methods when the network is large. When the network is small we use projection and back-substitution to solve for f_{con}, f_{rot} and ϕ. In many cases θ is not required, however when it is needed we solve $C^{\intercal}\theta = f_{\text{rot}}$ for θ.

2.3.2 Path Integrals

So far, the scalar potential ϕ has been characterized as the solution to a least squares problem, and the solution to a discrete Poisson equation. Here we derive an alternative characterization which will be useful when comparing the potential defined by the HHD to other potential decompositions.

Suppose we wanted to find the difference in the scalar potential at a and b. If the edge flow is conservative then solving for $\phi_b - \phi_a$ given f is easy. Pick a path connecting a to b. Then, define a vector y with E entries, such that $y_i = 0$ if the i^{th} edge is not included in the path, $y_i = 1$ if the path traverses the i^{th} edge in its positive direction, and $y_i = -1$ if the path traverses the edge in the negative direction. Then, the path "integral" of f over y is

simply $y^\intercal f$. By analogy with classical mechanics we say that the value of the path integral is the work to move from a to b over the path y. This analogy will be realized more fully in Section 6.5 where it is shown that, for appropriate physical systems $y^T f$ is the work to move from a to b over the path y (cf. [4, 5]).

If the edge flow is conservative then path integrals over f are path independent, so:

$$y^\intercal f = \phi_b - \phi_a. \tag{2.32}$$

Now suppose that the edge flow is not conservative. Then the path integral does not equal $\phi_b - \phi_a$, since path integrals are no longer path independent. Breaking f into its conservative and rotational parts we find:

$$y^\intercal f = y^\intercal (f_{\text{con}} + f_{\text{rot}}) = (\phi_b - \phi_a) + y^\intercal f_{\text{rot}}. \tag{2.33}$$

So, if the network does not obey detailed balance we cannot recover $\phi_b - \phi_a$ directly from a single path integral $y^\intercal f$ since any path integral may include a path dependent term $y^\intercal f_{\text{rot}}$ associated with the rotational field.

Now suppose that instead of computing one path integral we compute n path integrals, each over a different path $y^{(j)}$. Let Y be the $n \times E$ matrix whose columns are each a path from a to b. Then the ensemble of path integrals is given by $Y^\intercal f$:

$$Y^\intercal f = (\phi_b - \phi_a) + Y^\intercal f_{\text{rot}}.$$

Pick a weight vector w with n entries, with $\sum_i w_i \neq 0$. Then:

$$w^\intercal Y^\intercal f = \sum_{i=1}^{n} w_i(\phi_b - \phi_a) + w_i[{y^{(i)}}^\intercal f_{\text{rot}}].$$

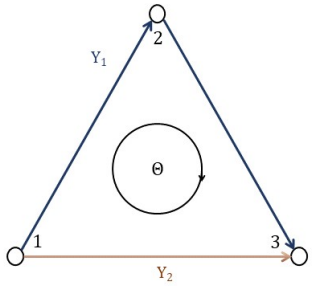

Figure 2.5: A pair of paths on a triangle network whose average path integral (with weights $1/3, 2/3$) is independent of the rotational field.

Or:

$$\frac{1}{\sum_i w_i} w^\intercal Y^\intercal f = (\phi_b - \phi_a) + \frac{1}{\sum_i w_i} w^\intercal Y^\intercal f_{\text{rot}}. \tag{2.34}$$

If w_i is chosen so that the second term cancels then the weighted average $\frac{1}{\sum_i w_i} w^\intercal Y^\intercal f = \phi_b - \phi_a$. The rotational flow, f_{rot}, is generally not known a priori, so we need to find a ensemble of paths Y and weights w such that $w^\intercal Y^\intercal f_{\text{rot}} = 0$ for any f_{rot}.

In some simple cases it is easy to guess an appropriate ensemble of paths Y and weights w. For example, given a triangle with nodes $1, 2, 3$, if we pick the paths $1 \to 2 \to 3$ and $1 \to 3$ with weights $1/3$ and $2/3$ respectively then:

$$w^\intercal Y^\intercal f_{\text{rot}} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} f_{\text{rot}} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} f_{\text{rot}}.$$

Since a triangle only has one loop $f_{\text{rot}} = \theta[1; 1; 1]$ for some choice of rotational potential θ. Therefore:

$$w^\intercal Y^\intercal f_{\text{rot}} = \frac{1}{3}\theta + \frac{1}{3}\theta - \frac{2}{3}\theta = 0$$

for any choice of θ (see Figure 2.5).

For some symmetric networks it is possible to pick Y and w by inspection. For example,

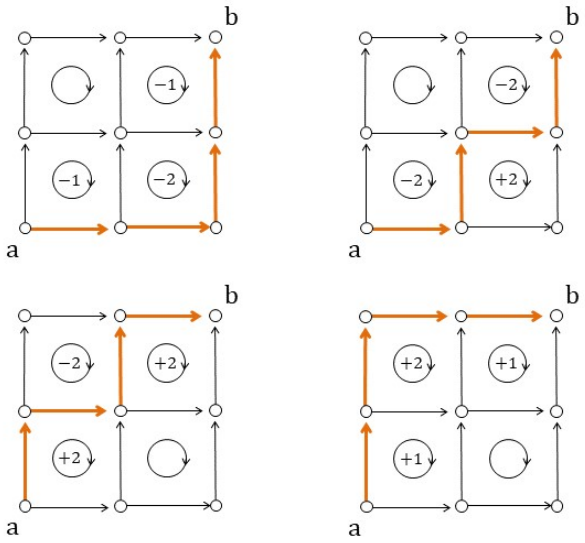

Figure 2.6: An ensemble of paths from a to b on a lattice whose average path integral (with weights all set to one) is independent of the rotational field. The contribution of the rotational potential on each loop to each path is shown inside of each loop. Note that summing over all of the paths cancels the contribution from every loop.

given a two dimensional lattice with equal side lengths, and given a at the bottom left corner and b at the upper right, then we should pick the paths as shown in Figure 2.6 and set all the weights to one.

The key to both these examples is that for every loop, the ensemble of paths includes at least one path that move around the loop in the clockwise direction and at least one path that moves around the loop in the counter-clockwise directions. Then, by picking the appropriate weights, we ensure that any loop's contribution to the average path integral is zero.

To translate from a geometric problem to an algebraic problem, note that any rotational edge flow can be written as the curl transpose of a rotational potential. Therefore, in order

for the rotational component of the average path integral to cancel we need:

$$w^\intercal Y^\intercal C^\intercal \theta = 0$$

for any choice of θ. This requires $CYw = 0$, or:

$$Yw \in \text{null}\{C\} \tag{2.35}$$

We can now restate our original problem. Given a and b we are looking for an ensemble of paths Y and weights w such that Yw is in the null space of the curl C, and $\sum_i w_i \neq 0$. If it is possible to find such a Y, w then:

$$\frac{1}{\sum_i w_i} w^\intercal Y^\intercal f = \phi_b - \phi_a. \tag{2.36}$$

The key questions are whether or not such Y, w exist, and how to find them.

In order to answer these questions, it is helpful to think a little more abstractly about the matrix Y. The matrix Y has n columns, each with E entries. That is, Y maps from a list of n objects to the space of edge flows. Therefore, if each column of Y is independent, the operation Yw takes a set of weights, and maps them to an n dimensional subspace of the space of edge flows. Notice that Y has many of the same properties as G and $C^\intercal$. Both G and $C^\intercal$ map from a set of objects (scalar and rotational potentials respectively) to subspaces of the space of edge flows. On the other hand, we originally introduced Y in order to evaluate path integrals. We wrote path integrals as products with $Y^\intercal$. That is, each element of $Y^\intercal f$ is a sum of f over a specific subset of edges. Similarly, each element of the divergence $G^\intercal f$ is a sum of f over the subset of edges neighboring a given node, and each element of the curl Cf is a sum of f over the subset of edges around a given loop. Thus

Y is similar to the operators G and C^T in that its columns map to a space of edge flows, and inner products with its columns evaluate sums over sets of edges. More pointedly, the range of Y is a space of edge flows that start at a and arrive at b.

Asking whether w exists such that $Yw \in \text{null}\{C\}$ is the same as asking whether or not the range of Y intersects the nullspace of C. If:

$$\text{range}\{Y\} \cap \text{null}\{C\} \neq \emptyset$$

then there must exist a nonzero vector w such that $w^\intercal Y^\intercal f_{\text{rot}} = 0$. If the range of Y is not contained in the range of $C^\intercal$ then the range of Y must intersect the null space of C since $\mathbb{R}^E = \text{range}\{C^\intercal\} \oplus \text{null}\{C\}$. In that case there exists a set of weights such that either $w^\intercal Y^\intercal f = \phi_b - \phi_a$ or $\sum_j w_j = 0$. If the dimension of the range of Y is greater than the dimension of the range of $C^\intercal$ then there is no way that the range of Y is contained in the range of $C^\intercal$. Therefore, if:

$$\dim[\text{range}\{Y\}] > \dim[\text{range}\{C^\intercal\}] = L \tag{2.37}$$

then there exists w such that either $w^\intercal Y^\intercal f = \phi_b - \phi_a$ or $\sum_j w_j = 0$.

Recall that when we defined Y the number of paths n was totally arbitrary. As long as each new path is independent of the previous paths (cannot be constructed by linear combination of the previous paths) then $\dim[\text{range}\{Y\}] = n$. Thus it remains to show that, for any a, b there exist a set of n independent paths from a to b such that $n > L$.

Notice that if we consider two distinct paths from a to b then the set of edges of the paths that are not shared by the two paths must form a loop. Moreover, if we consider a set of paths all starting at a and ending at b, then we can extract a set of loops from the set of paths (see Figure 2.7). As a basic rule we start by defining a base path from a to b.

Then for each additional path j added to Y we define a loop consisting of all the edges in y_1 not in y_j and all the edges in y_j not in y_1. We will refer to y_1 as the inner component of each loop and y_j as the outer component. Linear combinations of these loops have inner components restricted to the path y_1 and outer components restricted to the range of linear combinations of the paths where they differ from y_1. It follows that if each new path is independent then each new loop is independent of the previous loops. That is, everytime we introduce a new path the dimension of the subspace of loops spanned by the associated set of loops increases by one. If we introduce $L+1$ independent paths, then the set of loops defined by Y is a loop basis for the L dimensional space of loops.

This tight relationship between an ensemble of $L+1$ independent paths and a loop basis provides a straightforward method for constructing an ensemble of $L+1$ independent paths.

Recall that we defined a loop basis by first defining a spanning tree, then associating each loop in the loop basis with an edge left out of the spanning tree. The goal is to use the same scheme to construct a set of $L+1$ independent paths corresponding to the loop basis given by the spanning tree.

Pick a spanning tree. Then there is a unique path in the spanning tree from a to b. This is y_1 and will act as the inner component of each loop in the loop basis.

Consider paths one at a time that use exactly one chord to get from a to b. For convenience, order the chords (edges of the original graph left out of the spanning tree) so that the $j-1^{st}$ path corresponds to the j^{th} missing edge. If the path crosses an edge twice in opposite directions, then remove that edge from the path. Flip the direction of traversal of each edge until the path points monotonically forward from a to b.

After repeating this process for L missing edges, we have an ensemble of $L+1$ paths. These are all independent since all the paths (excluding the first) traverse an edge that none of the other paths traverses. It follows that there is no linear combination of the outer

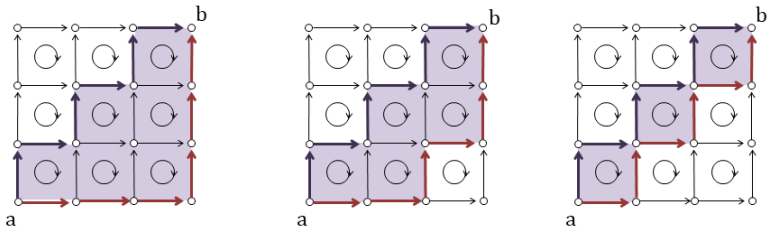

Figure 2.7: Three basis loops (shaded purple) formed from four paths from a to b. The inner path is shown in dark purple while the outer paths are shown in red.

paths (paths 2 to $L+1$) that is retricted to the inner path. Therefore there is no linear combination of all $L+1$ paths that is zero on every edge. It follows that Y has a trivial nullspace ($Yw = 0$ if and only if $w = 0$) and all of its columns are independent.

Then, by construction Y has $L+1$ columns and:

$$\dim[\text{range}\{Y\}] = L + 1 > L. \tag{2.38}$$

Therefore there must exist a set of weights w such that either the average path integral $w^\intercal Y^\intercal f = \phi_b - \phi_a$, or equals zero. In order to solve for w given Y we need only row reduce the matrix CY, and solve for its nullspace. As long as this nullspace includes a vector that is not orthogonal to the vector of all ones, $\mathbf{1}$, then there exists a w such that $w^\intercal Y^\intercal f = \phi_a - \phi_b$. Note that to interpret this product as a weighted average also requires that the nullspace of CY includes a vector with all nonnegative entries.

This reasoning suggests that, for any closed graph, the difference in scalar potential between two nodes is equivalent to the average work it takes to move between those nodes, albeit, given a set of paths Y with $\dim[\text{range}\{Y\}] > F$ and a set of weights nonnegative w such that $CYw = 0$.

While Y and w are not unique, the flow Yw is unique if $\sum_j w_j \neq 0$. To see this,

evaluate the divergence, $-G^\intercal Yw$. By assumption, every column of Y is a continuous path from node a to node b. Consider a specific path, say y_j. For any node, i, in the interior of the path, $w_j y_j$ has no divergence since $w_j y_j$ introduces one edge with flow w_j into i and one edge with flow w_j out of i. It follows that Yw has no divergence at any nodes except a and b. Every path in Y starts at a so the divergence of Yw at a is $\sum_j w_j$ which, if nonzero, can be set to one by convention. Similarly every path in Y ends at b so the divergence of Yw at b is $-\sum_j w_j = -1$. Therefore:

$$[-G^\intercal Yw]_i = \begin{cases} 1 \text{ if } i = a \\ -1 \text{ if } i = b \\ 0 \text{ else} \end{cases} \tag{2.39}$$

This shows that the divergence of Yw is uniquely defined when w can be chosen such that $\sum_i w_i = 1$. If $CYw = 0$ then Yw is curl free, so Yw is conservative. It follows that there exists some potential function u such that $-Gu = Yw$ and, u is uniquely defined (up to the addition of a constant) by the Poisson equation:

$$-G^\intercal Gu = G^\intercal Yw. \tag{2.40}$$

Since $G^\intercal Yw$ is unique so is Gu, and, since $Yw = -Gu$, so is the flow Yw. Therefore, for any choice of paths Y from a to b and weights w such that $\sum_j w_j = 1$ that satisfy $CYw = 0$, the flow Yw is unique.

To recover this particular flow let e_j be the j^{th} column of a $V \times V$ identity matrix. Then we can write $G^\intercal Yw = e_a - e_b$. It follows that u is given by a difference of two Green's

functions associated with the node Laplacian $G^\intercal G$:

$$u = -\left[G^\intercal G\right]^\dagger (e_a - e_b) \tag{2.41}$$

and:

$$Yw = G\left[G^\intercal G\right]^\dagger (e_b - e_a), \tag{2.42}$$

where † denotes the psuedoinverse associated with setting the average value of u to zero.

To make sense of this result note that, since $CYw = 0$:

$$(e_a - e_b)^\intercal \phi = (e_a - e_b)^\intercal \left[G^\intercal G\right]^{-1} G^\intercal f = \left[G\left[G^\intercal G\right]^{-1} (e_b - e_a)\right]^\intercal f = [Yw]^\intercal f$$

where the second equality is given by the discrete Poisson equation for the potentials (Equation (2.26)). So far we have considered the problem for arbitrarily chosen ensembles of paths. Now we will consider a special ensemble of paths that leads to a simple interpretation of the potential ϕ.

A simple, or unweighted, random walk on a network is a Markov process where the probability of going from node i to a neighboring node $j \in \mathcal{N}_i$ does not depend on which neighbor j is chosen. Start the walker at node a. Then, to update the state of the walker, draw a node uniformly from the neighbors of a. Then repeat this process, drawing uniformly from the neighbors of the current node, until the walker reaches b and exits the network.

The simple random walk defines a sampling scheme for drawing random trajectories from a to b. Given a trajectory y let m_j be the number of times the trajectory visits node j.

Then the probability of y is proportional to:

$$P(y) \propto \prod_{j \neq b} \left(\frac{1}{|\mathcal{N}_j|} \right)^{m_j} \tag{2.43}$$

with a normalizing constant associated with the condition of starting at a and ending at b.

If Y is the ensemble of all such paths, then the expected number of traversals of each edge in the positive direction is given by:

$$J = \sum_{y \in Y} P(y) y. \tag{2.44}$$

The average number of traversals, J, is equivalent to the current over each edge if we assign each edge resistance one, introduce a unit current at node a, and remove a unit current at node b [28].

By Kirchoff's laws, if we introduce a unit current at node a and remove a unit current at node b, then the divergence of J must be 1 at a, -1 at b, and zero everywhere else. That is:

$$-G^\intercal J = e_a - e_b. \tag{2.45}$$

By Ohm's law, the current across any edge is $J_{ij} = \frac{V_i - V_j}{R_{ij}}$, however we assumed that $R_{ij} = 1$ for all connected i, j so $J = -GV$. Which implies that J is curl free since $CG = 0$. These are precisely the conditions that uniquely defined Yw for any ensemble of paths Y and weights w so that $w^\intercal Y^\intercal f = \phi_b - \phi_a$. That is, $J = Yw$. It follows that if we let $\mathcal{Y}$ be the set of all possible trajectories from a to b and set $w(y \in \mathcal{Y}) = P(y)$ then:

$$\phi_b - \phi_a = \sum_{y \in \mathcal{Y}} P(y) y^\intercal f. \tag{2.46}$$

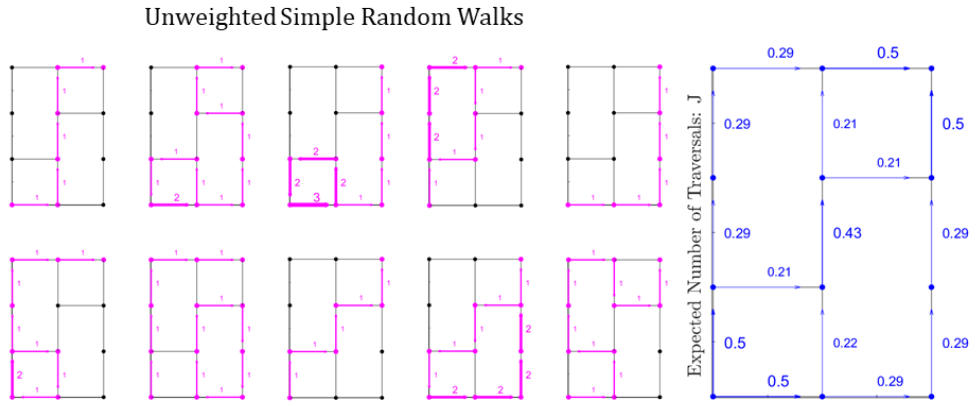

Figure 2.8: Ten sample trajectories are shown in magenta from $(0, 0)$ to $(3, 2)$, each sampled according to an unweighted random walk. The numbers next to each edge correspond to the number of positive traversals. The larger panel on the right shows the mean flux J over each edge given 10^6 sample trajectories. Note that, up to rounding error, the divergence of J shown in the right hand panel is zero at all nodes except the bottom left hand corner (a) and the upper right hand corner (b). Also note that, up to rounding error, the curl on each loop is zero.

Theorem 10 (Path Integral Interpretation)**.** *The difference in the scalar potential $\phi_b - \phi_a$ between any pair of nodes a, b is the expected value of the work it takes to move against the edge flow f over randomly drawn trajectories from a to b, where the trajectories are realizations of simple random walks on the graph $\mathcal{G}$ that start at a and end at b.*

Theorem 10 leads immediately to a Monte-Carlo method for computing the potential ϕ. First draw an ensemble of simple random walks starting at node a. Then evaluate the path integral over each random walk, and average the results. By the law of large numbers, the average value of the path integral will converge to its expected value, which, by Theorem 10, is exactly the value of its scalar potential. An example is illustrated in Figure 2.8, and convergence of the sampled flux $J(n)$ to the flux J is shown in Figure 2.9.

Since the potentials are a linear function of the edge flows f, Theorem 10 generalizes naturally to sets of nodes. Let A and B be sets of nodes. Let $\bar{\phi}(A)$ and $\bar{\phi}(B)$ be the average

potential on each set. Then $\bar{\phi}(B) - \bar{\phi}(A)$ is exactly the expected work to move from a node a drawn randomly from A to a node b drawn randomly from B over a randomly drawn path. If we assign a probability of drawing a from A that is not uniform and a probability of drawing b from B that is not uniform then the expected value of the work is equal to the difference in weighted average of the potentials between the two sets.

Let π_a be the probability of drawing a and π_b be the probability of drawing b. By assumption $\sum_{a \in A} \pi_a = 1$ and $\sum_{b \in B} \pi_a = 1$. If a and b are drawn independently then the joint probability of drawing any pair a, b is just $\pi_a \pi_b$. Then the expected value of the work W to move from A to B where a, b are drawn randomly is:

$$\mathbb{E}[W] = \sum_{a \in A, b \in B} \pi_a \pi_b (\phi_b - \phi_a) = \sum_{b \in B} \pi_b \phi_b \sum_{a \in A} \pi_a - \sum_{a \in A} \pi_a \phi_a \sum_{b \in B} \pi_b = \mathbb{E}[\phi_b] - \mathbb{E}[\phi_a].$$

It follows immediately that the difference in weighted average between the potentials in any two sets of nodes (with positive weights) is the expected value of the work it takes to move between the two sets over randomly drawn trajectories that connect the two sets.

Corollary 10.1. *Suppose A and B are both sets of nodes, π_a is the probability of drawing a node a from set A, and π_b is the probability of drawing a node b from the set B. Define the expected potentials $\bar{\phi}(A) = \mathbb{E}[\phi_a], \bar{\phi}(B) = \mathbb{E}[\phi_b]$ where a and b are drawn randomly from A and B. Then the expected work to move from A to B, with randomly drawn endpoints a, b, and trajectories drawn from a simple random walk, is the difference in the expected potentials.*

These results generalize naturally to the rotational potential θ if the network $\mathcal{G}$ is planar, using the usual duality argument.

Theorem 10 provides an alternative way to think about the scalar potential associated

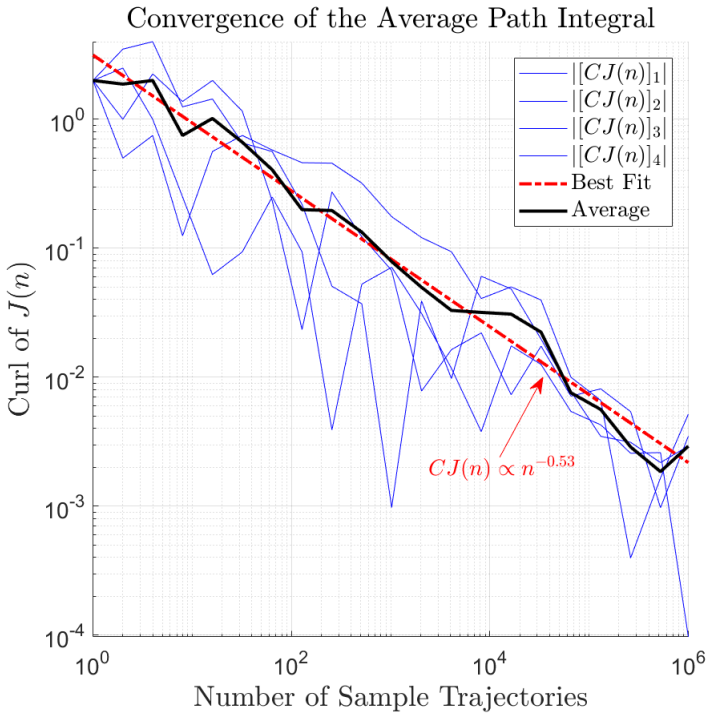

Figure 2.9: Convergence of the Monte Carlo approximation, $J(n)$, to the flux J whose curl is zero for the example shown in Figure 2.8. Blue lines represent the curl on each basis loop. The thick black line represents the average of the blue lines. The red dotted line represents the best fit line to the average curl. The slope of the red line is -0.53, which matches the expected convergence rate $\mathcal{O}(n^{-1/2})$. As the curl on each loop vanishes the contribution of the rotational potential on each loop to the average path integral vanishes.

with the HHD. In Section 7.3.1 and Section 7.3.2 we will show that the HHD is the appropriate potential framework for studying Markov process that are diffusion dominated since in a purely diffusive limit the Markov process converges to a weighted random walk. In Section 2.4.2 we generalize Theorem 10 to weighted random walks.

2.4 Generalization

2.4.1 Extended Cycle Bases

In some cases the requirement that the set of cycles used to define the curl is a cycle basis is too restrictive. For example, on a complete network the set of all triangles is a natural collection of cycles, but is not a cycle basis since it includes too many cycles. There are V choose 3 distinct triangles in a complete graph, but only $(V-2)(V-1)/2$ loops in its cycle basis. This can be seen by noting that a complete graph has V choose 2 distinct edges, so $L = E - (V-1) = V(V-1)/2 - (V-1) = (V-1)(V-2)/2$. In contrast V choose 3 equals $V(V-1)(V-2)/6$ which is $V/3$ times larger than L.

A cycle basis for a complete graph can be formed by picking an initial node, then considering all triangles involving that node. This equals the fundamental cycle basis if the spanning tree is chosen to be a star centered at the initial node. Clearly this choice of basis is arbitrary since there are V possible initial nodes. While the choice of cycle basis does not change the component flows, or the scalar potential, it does change the rotational potential θ since θ is defined on the chosen cycle basis. Instead of representing rotation on all triangles involving a particular node of the graph we may prefer to represent rotation on a different set of triangles.

A similar problem arises in non-planar square lattices. The set of all squares is a natural set of cycles, but includes too many cycles to be a cycle basis [32]. An extended discussion of cycle bases on lattices is presented in Section 3.3, and the complexity of constructing a cycle basis on lattices is a good motivation for considering extended sets of cycles.

Suppose that $\mathcal{C}$ is extended to include all cycles of a certain kind, such that a subset of the cycles form a cycle basis. As noted before, this does not change the rotational component since the range of the C^{T} is unchanged by adding additional cycles to a cycle

basis. However, after adding extra cycles, $C^\intercal$ no longer has a trivial nullspace. It follows that $C^\intercal\theta = f_{\text{rot}}$ will not have a unique solution.

The rotational potential on a set of cycles that includes a cycle basis can be defined uniquely by introducing an additional requirement on θ. This requirement should be chosen to promote a desired feature in the decomposition. A natural requirement is to minimize $||\theta||_0$ or $||\theta||_1$ over all θ such that $C^\intercal\theta = f_{\text{rot}}$. Then θ is defined by:

$$\theta = \text{argmin}_{v\in\mathbb{R}^{|\mathcal{C}|}}\{||v||_p | C^\intercal v = f_{\text{rot}}\}. \tag{2.47}$$

where $p = 0$ or 1. Lim and Jiang advocate for $p = 1$ [16], since setting $p = 1$ ensures that θ is the most parsimonious representation of rotation on the chosen set of cycles, and under regularity conditions on C may also be the sparsest solution on the expanded cycle basis even when f is perturbed by noise [52, 53]. The $p = 1$ problem can be solved efficiently since it can be framed as a linear programming problem. The requirement $C^\intercal\theta = f_{\text{rot}}$ fixes f to the rotational subspace. Then define the vector $\hat{v}$ with $2|\mathcal{C}|$ entries. Let $\hat{v}_h = v_h$ if $v_h > 0$ and 0 if $v_h < 0$, and let $\hat{v}_{2h} = -v_h$ if $v_h < 0$ and 0 if $v_h > 0$. Then solving for θ with $p = 1$ is the same as minimizing $\sum_{h=1}^{2\mathcal{C}} \hat{v}_h$ with the constraint $\hat{v}_h \geq 0$. This is a minimization problem over a convex polytope defined by the intersection of the rotational subspace $[C^\intercal; -C^\intercal]\hat{v} = f_{\text{rot}}$ with the nonnegativity constraint with a linear objective function. It follows that, if the problem has a unique solution, then the solution is a vertex of the polytope defined by the intersection of the subspace $C^\intercal\theta = f_{\text{rot}}$ with the nonnegativity constraint. If the solution is not unique then it can be expressed as a convex combination of the vertices. Each vertex has at most L nonzero entries.

When solving for θ on complete graphs we adopt this approach. In general we find that the solution has, as expected, exactly L nonzero entries, so the l_1 optimization problem

chooses a set of L triangles forming a cycle basis from the set of all possible triangles. This is an example of cycle basis pursuit, in which the cycles in the cycle basis are ultimately chosen from a larger set of cycles in order to give the most parsimonious representation of the rotational flow possible. It is important to note that the l_1 solution may not be stable under perturbations of the edge flow. In those situations alternative sparse solvers can be used. If the edge flow is perturbed by noise, or can only be estimated from data then there is a posterior distribution of possible edge flows and a Bayesian approach could be used (cf. [54, 55])

2.4.2 Weights

The HHD can also be generalized by introducing a set of weights W to the familiar HHD equation $-G\phi + C^\intercal\theta = f$ [16]. Throughout this section W will denote a diagonal matrix with positive diagonal entries.

In principle there are five different places we could introduce weights: before G, before ϕ, before $C^\intercal$, before θ, or before f. However, most of these choices only lead to trivial modifications of our existing theory. For example, suppose we introduced weights W_1, W_2 and W_3 so that:

$$-GW_1\phi + C^\intercal W_2\theta = W_3 f.$$

Then, if we define $\tilde{\phi} = W_1\phi$, $\tilde{\theta} = W_2\theta$, and $\tilde{f} = W_3 f$ then $-G\tilde{\phi} + C^\intercal\tilde{\theta} = \tilde{f}$ which is the standard HHD equation for scaled potentials $\tilde{\phi}$ and $\tilde{\theta}$ given the scaled edge flows $\tilde{f}$. Therefore there is no need to treat these weightings as a fundamental generalization of our existing theory.

This leaves the weighting:

$$-W_1 G\phi + W_2 C^\intercal \theta = f. \tag{2.48}$$

Notice that if $W_1 = W_2 = W$, and W is invertible, then we could multiply across by W^{-1} and would be back in the previous case with $\tilde{f} = W^{-1} f$. Therefore the only case with invertible weights that modifies our existing theory is the case $W_1 \neq W_2$.

This begs the question, given invertible $W_1 \neq W_2$ do the potentials ϕ, θ exist for any f, and if the network is closed, are they unique?

Theorem 11 (The Weighted HHD)**.** *For any finite network with closed boundaries, and any invertible edge weights W_1, W_2, there exist a unique pair of potentials (up to the addition of a constant) satisfying:*

$$-W_1 G^\intercal \phi + W_2 C^\intercal \theta = f$$

where ϕ, θ are the least squares solutions to the weighted Poisson equations:

$$\begin{aligned} G W_1 W_2^{-1} G\phi &= -G^\intercal W_2^{-1} f \\ C W_1^{-1} W_2 C^\intercal \theta &= C W_1^{-1} f. \end{aligned} \tag{2.49}$$

Proof. To prove Theorem 11 we start by introducing a special case. Suppose $W_1 = W_2^{-1}$. Then define the weighted operators $\tilde{G} = W_1 G$ and $\tilde{C}^\intercal = W_2 C^\intercal$. It is trivial to see that the weighted operators are still orthogonal since:

$$\tilde{C}\tilde{G} = C W_2^\intercal W_1 G = C W_1^{-1} W_1 G = CG = 0.$$

Therefore the Fundamental Theorem of Linear Algebra guarantees the existence of ϕ

and θ. To ensure uniqueness we need the joint nullspace of the weighted divergence and weighted curl to be empty. Equivalently, we need to show that the range of the weighted gradient is equivalent to the nullspace of the weighted curl. Suppose z is in the nullspace of C. Then $W_1 z$ is in the nullspace of $\tilde{C}$. Therefore null$\{\tilde{C}\} = W_1$null$\{C\}$. In a finite network null$\{C\} =$ range$\{G\}$. By definition $\tilde{G} = W_1 G$ so range$\{\tilde{G}\} = W_1$range$\{G\}$. Therefore:

$$\text{null}\{\tilde{C}\} = W_1\text{null}\{C\} = W_1\text{range}\{G\} = \text{range}\{\tilde{G}\}.$$

This proves existence and uniqueness in the special case $W_2 = W_1^{-1}$. To solve for the potentials we solve the least squared equations associated with the weighted operators. That is, we multiply the HHD equation by $\tilde{G}^\intercal$ and $\tilde{C}$ to get:

$$\begin{aligned} G^\intercal W_1^2 G\phi &= -G^\intercal W_1 f \\ C W_1^{-2} C^\intercal \theta &= C W_1^{-1} f. \end{aligned} \tag{2.50}$$

Now suppose $W_2 \neq W_1^{-1}$. Then we can generalize from the special case by symmetrizing the problem. Given:

$$-W_1 G\phi + W_2 C^\intercal \theta = f$$

multiply both sides by $W_1^{-\frac{1}{2}} W_2^{-\frac{1}{2}}$. Since the weights are diagonal they commute, therefore this product can be written:

$$-W_1^{\frac{1}{2}} W_2^{-\frac{1}{2}} G\phi + W_1^{-\frac{1}{2}} W_2^{\frac{1}{2}} C^\intercal \theta = W_1^{-\frac{1}{2}} W_2^{-\frac{1}{2}} f.$$

Now the weight in front of the gradient is the inverse of the weight in front of the adjoint curl. That is, by symmetrizing the problem we arrive in the special case where the two weights are each other's inverse. It follows that if the weights are invertible the potentials

exist and are unique. Moreover, by substituting into Equation (2.50) the potentials satisfy the corresponding Poisson equations:

$$\begin{aligned} G^\intercal W_1 W_2^{-1} G\phi &= -G^\intercal W_2^{-1} f \\ C W_1^{-1} W_2 C^\intercal \theta &= C W_1^{-1} f. \end{aligned} \tag{2.51}$$

□

Theorem 11 plays the same role in our generalized theory as Theorem 5 played in defining the thermodynamic potentials ϕ, θ. Theorem 11 guarantees that the weighted HHD is well defined for all invertible weights.

Corollary 11.1. *Given the weighted HHD:*

$$-G\tilde{\phi} + W^{-1} C^\intercal \tilde{\theta} = f = -G\phi + C^\intercal \theta$$

the generalized rotational potentials are related to the unweighted rotational potential by the change of weights formula:

$$C W^{-1} C^\intercal \tilde{\theta} = C C^\intercal \theta. \tag{2.52}$$

Alternatively given the weighted HHD:

$$-WG\tilde{\phi} + C^\intercal \tilde{\theta} = f = -G\phi + C^\intercal \theta$$

the generalized scalar potential is related to the unweighted scalar potential by the change of weights formula:

$$G^\intercal W G \tilde{\phi} = G G^\intercal \phi. \tag{2.53}$$

Proof. The proof is trivial. Multiply either equation by C or $G^\intercal$ in order to find the relation between the generalized potential and unweighted potential. □

If we pick a convention for the constant associated with the scalar potential then both Laplacians are invertible. Either of the change of weights formulas can then be used to derive the weighted potentials from the unweighted potentials. The change of weights equation will appear again when we study the weak rotation limit of Markov processes on networks in Section 7.3.2.

The introduction of weights also leads to a more flexible characterization of the potentials using energy norms. In general, if a matrix A is positive definite, then the associated energy norm is $||v||_A^2 = v^\intercal A v$. When the HHD is unweighted, the potentials minimize an l_2 norm of the error left over when approximating the flow using the potentials. When weighted, the generalized potentials minimize the energy norm of the error.

Corollary 11.2. *Given a weighted HHD equation:*

$$-G\phi + W^{-1}C^\intercal\theta = f \tag{2.54}$$

the scalar potential ϕ is the unique minimizer:

$$\phi = \text{argmin}_u\{||Gu + f||_W\} \tag{2.55}$$

and the rotational potential θ is the unique minimizer:

$$\theta = \text{argmin}_v\{||C^\intercal v + Wf||_{W^{-1}}\}. \tag{2.56}$$

Proof. If ϕ minimizes $||Gu + f||_W$ then it minimizes $||Gu + f||_W^2$. By definition:

$$||Gu + f||_W^2 = (Gu + f)^\intercal W(Gu + f) = u^\intercal G^\intercal W G u + 2f^\intercal W G u + f^\intercal W f.$$

Therefore:

$$\partial_u ||Gu + f||_W^2 = 2G^\intercal W G u + 2f^\intercal W G.$$

Setting the derivative to zero recovers the weighted Poisson equation:

$$G^\intercal W G u = -G^\intercal W f.$$

But, from Theorem 11, ϕ is the unique solution to the weighted Poisson equation so $u = \phi$. This must be the global minimum since $||Gu + f||_W^2$ is a convex function. The proof of the second relation involving θ follows analogously. □

Corollary 11.2 provides an alternative interpretation of the generalized potentials. Each potential is defined by minimizing the energy norm of the error in approximating f with $-G\phi$ or $W^{-1}C^\intercal\theta$. Notice that the scalar potential is biased to give an accurate approximation of f on edges where the weights are large, and the curl on edges where the weights are small.

Combined Theorem 11, Corollary 11.1, and Corollary 11.2 complete the basic algebraic description of the generalized potentials. Our next goal is to show that the path integral interpretation (Theorem 10) can be generalized to account for weights. The generalization depends on using a continuous time Markov chain to sample trajectories. See [56] or Chapter 6 for a definition of continuous time Markov chains.

Theorem 12 (Path Integral Interpretation for the Weighted HHD)**.** *Given a finite network with closed boundaries, and an invertible pair of weights W_1, W_2 with corresponding weighted HHD:*

$$-W_1 G\phi + W_2 C^\intercal \theta = f$$

the difference in scalar potential $\phi_a - \phi_b$ is equal to the expected value of the work evaluated against $\tilde{f} = W_1^{-1} f$ to move from a to b on paths y drawn from a continuous-time Markov process with instantaneous transition rates $W = W_2^{-1} W_1$ from $i(k)$ to $j(k)$ and $j(k)$ to $i(k)$.

Proof. First, multiply the weighted HHD on the left by W_2^{-1}. Then:

$$-W_2^{-1} W_1 G\phi + C^\intercal \theta = W_2^{-1} f.$$

Let $\tilde{f} = W_1^{-1} f$ and $W = W_2^{-1} W_1$. Then:

$$-WG\phi + C^\intercal \theta = W_2^{-1} f = W_2^{-1} W_1 \tilde{f} = W\tilde{f}. \tag{2.57}$$

By Theorem 11, the scalar potential, ϕ, is the unique solution to the weighted Poisson equation:

$$G^\intercal W G\phi = -G^\intercal W_2^{-1} f = -G^\intercal W \tilde{f}. \tag{2.58}$$

Next consider a continuous time Markov process with transition matrix $G^\intercal WG$. This is a symmetric transition matrix, with instantaneous transition rates from $i(k)$ to $j(k)$ and $j(k)$ to $i(k)$ equal to w_k. Continuous time Markov processes will be discussed in more detail in Section 6.2.

Then the paths y are sampled from the corresponding skeleton process, which records

only the sequence of states visited along a sample trajectory, not the transition times. The skeleton process is a discrete-time Markov process where the probability of moving to node j from node i is equal to $w_{ij}/(\sum_{l\in\mathcal{N}_i} w_{il})$.

Let $\mathcal{Y}$ be the ensemble of all possible paths from a to b, and let y be a vector representing the number of positive traversals of each edge on a path from a to b. Sample y from $\mathcal{Y}$ using the weighted random walk. Then, the expected number of positive traversals of each edge is equal to the steady state current between nodes a and b in an electrical network with conductances set to w_{ij} [28]. Denote this current $J = \mathbb{E}[y]$.

The steady state current must obey both of Kirchoff's laws. Therefore the divergence of J must satisfy $G^\intercal J = e_a - e_b$ where e_j is the j^{th} column of an identity matrix, and the curl of J must be zero, $CJ = 0$. It follows that J is equal to a weighted gradient of a voltage u of each node. From Ohm's law $J_{ij} = w_{ij}[u_j - u_i]$ so:

$$J = -WGu. \tag{2.59}$$

Therefore:

$$-G^\intercal WGu = G^\intercal J = e_a - e_b. \tag{2.60}$$

Multiplying across by the pseudo-inverse of the weighted Laplacian recovers u (up to a constant). Then, evaluating the gradient gives:

$$J = -WGu = -WG[G^\intercal WG]^{-1}(e_a - e_b). \tag{2.61}$$

Therefore the expected work to move from a to b against $\tilde{f}$ is:

$$\mathbb{E}[y^\intercal \tilde{f}] = J^\intercal f = -(e_a - e_b)[G^\intercal WG]^{-1}G^\intercal W\tilde{f}. \tag{2.62}$$

Define a new function on the nodes $\hat{u}$ such that $\hat{u}_a - \hat{u}_b = \mathbb{E}[y^\intercal \tilde{f}]$. Then:

$$\hat{u} = -[G^\intercal W G]^{-1} G^\intercal W \tilde{f}$$

or:

$$[G^\intercal W G]\hat{u} = -G^\intercal W \tilde{f}. \tag{2.63}$$

But this equation is exactly the same as the weighted Poisson equation that defined ϕ so $\phi = \hat{u}$, or:

$$\phi_a - \phi_b = \mathbb{E}[y^\intercal \tilde{f}] \tag{2.64}$$

where $\tilde{f} = W_1^{-1} f$ and the paths y are sampled from $\mathcal{Y}$ according to the weighted random walk with weights $W = W_1 W_2^{-1}$.

□

Theorem 12 provides a clear connection between the choice of weights in the HHD and the choice of paths along which we evaluate work between points. The difference in choice of weights between the gradient and the curl maps to weighting the random walk between nodes used to evaluate the potential.

Weighting allows us to treat the topology of the network in a more flexible way. When the HHD is unweighted, the operators G and C are defined by assuming all the edges are equivalent. Then J is a natural edge importance measure for passage from a to b in an unweighted graph. If we introduce $W_1 \neq W_2$ then J is still an edge importance measure for passage from a to b, but now in the context of a weighted graph with weights $W = W_1 W_2^{-1}$.

For example, suppose we want to measure work with respect to f, but over paths drawn from a simple random walk weighted by some W. Then simply set W_1 to the identity and $W_2 = W^{-1}$. Then the solution to either $-G\phi + W^{-1} C^\intercal \theta = f$ or $-WG\phi + C^\intercal \theta = Wf$

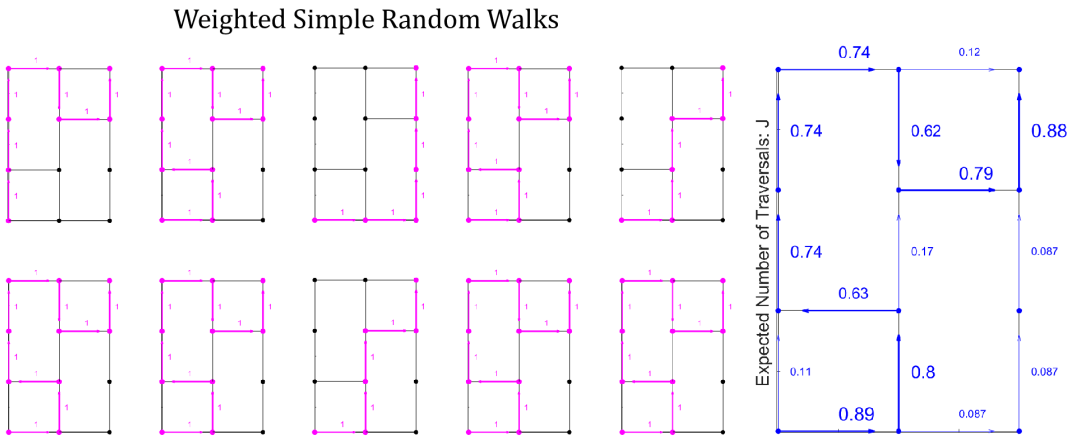

Figure 2.10: Ten sample trajectories are shown in magenta from $(0, 0)$ to $(3, 2)$ sampled according to a weighted random walk. The numbers next to each edge correspond to the number of positive traversals. The larger panel on the right shows the mean over 1000 sample trajectories. The weights were all set to either 1 or 20. The edges with weight 20 are the edges that appear with a large expected flux J. Compare this to Figure 2.8. Note that the divergence of the current J is zero (up to rounding errors) at all nodes in the graph except the bottom left and top right. Also notice that the current is clearly concentrated about a particular path.

satisfies $\phi_a - \phi_b = \mathbb{E}[y^\intercal f]$ with y sampled from a random walk weighted by W.

It immediately follows that every generalized potential is equivalent if the flow on the right hand side is conservative. If the flow is conservative then the work over a path is independent of the path taken, so the distribution of paths has no influence on the average work over the ensemble.

Corollary 12.1. *The generalized scalar potential ϕ defined by $-W_1 G\phi + W_2 C^\intercal \theta = f$ does not depend on W_2 if $W_1^{-1} f$ is curl free.*

In conclusion, by introducing weights to the HHD we define a family of generalized potentials. Provided the weights are invertible, the potentials exist, are unique (up to the addition of a constant), and satisfy weighted Poisson equations. Moreover, the corresponding scalar potentials are the best approximations to the edge flow f in the energy norm

associated with the weights, and the difference in scalar potential between points is given by the expected work to traverse between these points on a path drawn from a random walk with the symmetric transition matrix $G^\intercal WG$.

2.4.3 Networks with Open Boundaries

Thus far the harmonic component has only played a passing role in our theory. On a finite network, with an appropriately chosen curl, the harmonic component is always zero, so could be safely ignored. This is not true if the network has open boundaries. A network has an open boundary if there are edges that leave the network. Then the edge flow may flow in and out of the network through the boundary.

Open networks occur in a variety of systems for a variety of reasons. The domain may only be partially observable, or the topology may only be known on a subset of the full domain. Edge rates may only be knowable on a subset of the domain. More generally, the system may only be governed by a random walk process on a portion of the domain, or the random walker may be allowed to leave the domain. Even if the global network is available, global potentials may not be computable, or may not be of interest. Alternatively if the network is very large, or infinite, it may be necessary to study only a subset of the network.

This section will focus on the following open boundary problem: suppose the network of interest is a subset of a larger finite network with closed boundaries satisfying all the assumptions introduced in Section 2.2. Then there is no harmonic component on the full network and the potentials are uniquely defined. The goal is to solve for potentials on the subset. The subset is the domain of interest, the complement of the subset is the reservoir. Edges passing from the domain of interest to the reservoir are the boundary. A network is open if the domain of interest is connected to the reservoir.

The open boundary problem is more complicated than the closed boundary problem because the harmonic component on a finite network with open boundaries may be nonzero, even if there is no harmonic component on the full network. If it were possible to separate the edge flow into a unique conservative and rotational component then, restricting G to act only on nodes and edges in the domain of interest, and C to only act on loops in the domain of interest, the scalar potential and rotational potentials would be uniquely defined (up to a constant) by $-G\phi = f_{\text{rot}}$ and $C^{\intercal}\theta = f_{\text{con}}$. Unfortunately, a conservative or rotational flow on the full network passing through the subset may be harmonic on the subset (no divergence at any node in the subset or curl on any loop in the subset). Thus it is impossible to tell whether and what parts of a harmonic flow observed on the subset are conservative or rotational on the full network. As a result the potentials, which were uniquely defined for finite closed networks, are not uniquely defined for closed networks with open boundaries.

For example, suppose the larger network is a loop of nodes, and the domain of interest is some connected subset of nodes on that loop leaving at least two nodes in reservoir. Now imagine that f is a nonzero constant on the domain of interest and the edges passing into and out of the domain, and points in the same direction on every edge. Then the divergence of f is zero at every node in the domain since the flux into and out of every node is identically zero. The curl must also be zero since there are no loops in the domain of interest. Since f is both divergence and curl free it is harmonic. Examples of harmonic edge flows are shown in Figure 2.11 and Figure 2.12.

Whenever we focus on a subset of a larger network, we exclude information about f on the reservoir. The previous example could be generated by a scalar potential that decreases by a fixed amount on the sequence of nodes running from one side of the domain to the other, including the two nodes in the reservoir at either end. It is also possible that a fixed rotational potential on the loop drives the circulation. Without f on the full domain, it is

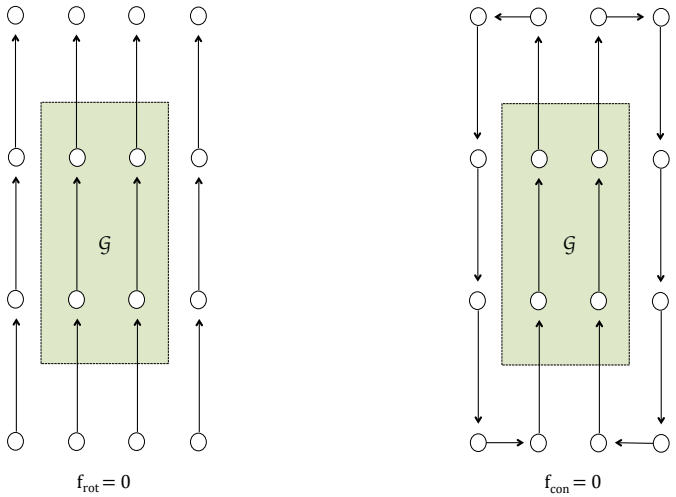

Figure 2.11: Two possible interpretations of the flow through an open network. The green shaded region is the domain of interest. The flow through the domain is both divergence and curl free so is harmonic on the domain of interest. The flow on the left is conservative on the full network, and the flow on the right is rotational, but both are identical on the domain of interest.

impossible to distinguish whether or not f on the domain is flowing from a node in the reservoir to another node in the reservoir, or flowing in a circuit. Therefore it is impossible to distinguish between rotational and conservative fields without f on the full network. As a result, it is impossible to decompose f into ϕ and θ uniquely given only the flow on the subset.

When a network has open boundaries, the HHD may include a nonzero harmonic component, as described in Theorem 3. The same problem arises in the continuum when there may be a flow through the boundary of the domain [8, 30]. In some cases it is sufficient to solve for a conservative, rotational, and harmonic component (cf. [16]), while in others we may desire a unique decomposition into only a conservative and rotational component. The latter is only possible if we either are given boundary conditions, are willing to make assumptions about the harmonic component, or define a convention for

handling the harmonic component. Any two decompositions of this kind differ only in how h is represented [30].

Dirichlet boundary conditions can be introduced if the network topology is known outside of the domain of interest. Then if the scalar potential is fixed at each node outside the domain but neighboring a vertex in the domain, and the rotational potential is fixed on each loop including some nodes inside, and some nodes outside the domain of interest. Neumann boundary conditions are often more natural, and assume that a certain component of the flow on boundary edges is conservative and the rest is rotational [8]. Common Neumann boundary assumptions include the assumption that all flow leaving the network is conservative, or the assumption that all flow along edges running between nodes on the boundary is rotational [30]. Under either condition the decomposition is still unique [31, 57]. Boundary constraints are the most widely used method for solving Poisson's equation in the continuum [8].

Unfortunately, not all problems lead unambiguously to natural assumptions about potentials or fields at the boundary. In that case, fixing boundary conditions may not give the best results, and in some cases may lead to serious boundary artifacts. Inappropriate boundary conditions can "create strong coupling between the component flows" $f_{\text{con}}, f_{\text{rot}}$ and the "shape and orientation of the boundary" [30]. This is generally undesirable. If the domain of interest is a subset of a larger network then the potentials should, ideally, not change if a node is added or removed from the boundary of the domain. That is, the computed potentials should be boundary independent, or at least close to boundary independent. If the computed potentials are not boundary independent then the decomposition may be called into question all-together.

One solution is to associate any harmonic component of f with potentials located at points in the reservoir (outside the domain of interest). In this context h is "external", and

is not included in either the conservative or rotational edge flow. These flows are considered "internal" since they are driven by potentials inside the domain of interest. The associated decomposition solves for ϕ and θ independent of h. This approach is the "natural HHD" proposed by Bhathia et al in [30].

The natural HHD is attractive for a number of reasons. First, it requires no a priori information about the potentials in the reservoir, or the flow on the boundary, since it is based on a convention regarding h rather than an assumption about the potentials. It also does not require boundary assumptions that guess potentials in the reservoir. In some cases the natural HHD has been shown to approximate fields computed on the full domain more accurately, and to introduce fewer boundary artifacts [30]. This is true when the Green's functions are sharply peaked, have bounded maxima, and the field f has little to no divergence or curl at the boundary.

To perform the natural HHD, compute the Green's functions for the full network, or a much larger subset of the network. Then compute the potentials within the domain of interest by summing the Green's functions for the larger network over the divergence/curl of f on only the nodes/loops inside the domain of interest [30]. This amounts to applying the inverse (or pseudoinverse) of the Laplacians to the divergence and curl of the flow on only the domain of interest. While conceptually straightforward, this may be inefficient numerically since it requires forming the inverse or psuedoinverse of the Laplacians on a larger network than the domain of interest.

It is generally advisable to decompose the network using a variety of reasonable constraints. It is also generally good practice to vary the boundary itself when considering a subset of a larger network. Variations in the computed potentials help determine how much the solutions depend on the chosen constraints and boundary. When it is not possible to vary the boundary then it may not be easy to determine how much of the computed solution

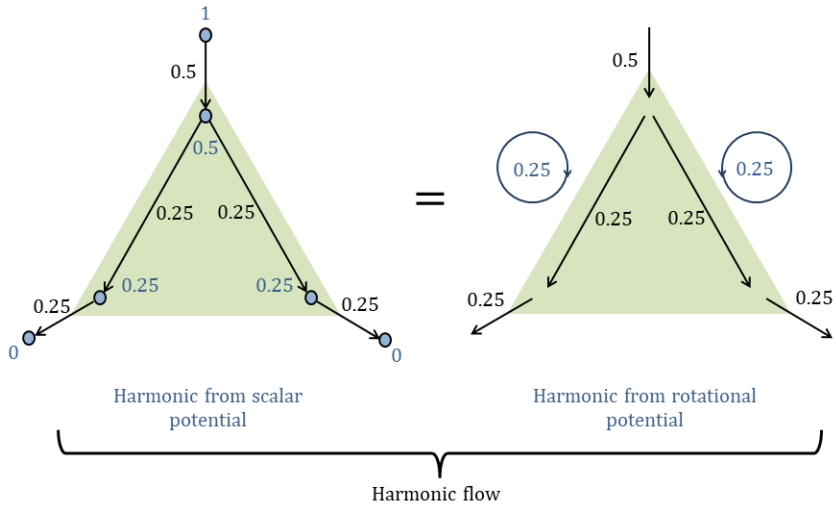

Figure 2.12: Two possible interpretations of the flow through an open network with specified edge flow. The green shaded region is the domain of interest. Black numbers denote the edge flow relative to the directions indicated by the arrows in both figures. The flow through the domain is both divergence and curl free, so is harmonic on the domain of interest. The flow on the left is conservative on the full network, and the flow on the right is rotational, but both are identical on the domain of interest. **Left:** Blue numerals indicate the scalar potential ϕ at the corresponding nodes (blue shaded circles). The rotational potential is zero. **Right:** Blue numerals indicate the rotational potential around the two loops in the network. The scalar potential is zero.

is "true" and how much is "artifact". Admittedly what is considered artifact is interpretive, and depends on the problem.

Regardless, it is essential to be aware of the assumptions introduced by the chosen constraint or boundary. If these assumptions are well motivated by the problem, then the computed potentials may be well justified, even if boundary dependent. If the assumptions introduced are not well motivated, then the computed potentials should be taken at face-value if and only if they are largely boundary independent. We advocate for first computing the harmonic component and the potentials under the assumption that the harmonic component is entirely external, and thus is a separate component of the flow not generated by either potential. Then boundary conditions can be introduced to assign the harmonic

component to either the rotational or conservative flows.

2.5 Summary

In this chapter we introduced the combinatorial Helmholtz-Hodge Decomposition. The decomposition separates an edge flow on a graph into two components. The first component is conservative and is the gradient of a scalar potential. The second is rotational, and is associated with the curl transpose of a rotational potential analogous to the vector potential used to describe magnetic fields. If the network is closed then the harmonic component is necessarily zero. If the network is open then there may be a harmonic flow entering and leaving the network through the boundary. Depending on chosen constraints this flow may be interpreted as separate from the rotational and conservative flows, or be decomposed itself into an assumed rotational and conservative component.

The HHD can be posed in a variety of ways. First, as projection onto a pair of orthogonal subspaces, second, as a pair of discrete Poisson equations, and third, as a pair of least-squares problems. Each of these characterizations is useful, as they lead to different solution methods and allow for alternative derivations of the HHD (see Chapter 4). We also showed that differences in the scalar potential at distinct vertices associated with the HHD can be interpreted as the average work over an ensemble of randomly drawn paths between those vertices. This path integral formulation provides an alternative perspective on the scalar potential that is useful for comparison to other potentials , and will help explain the utility of the HHD for analyzing the dynamics of Markov processes that are dominated by diffusion instead of drift.

The HHD can be generalized in two crucial ways. First, the choice of cycles used to define the curl is flexible, and it is possible to work with large cycle sets that contain cycle bases. Second, weights can be introduced to the HHD that give priority to the conservative

or rotational components on particular edges. We showed that, provided the weights are nonzero, then the potentials are still uniquely defined, satisfy weighted Poisson equations, are solutions to weighted least squares equations, and the difference in the scalar potential at distinct vertices is still the average work over randomly drawn paths. The weighted HHD is the essential tool for studying the limiting behavior of Markov processes that are diffusion dominated (see Section 7.3.1 and Section 7.3.2).

Chapter 3

Examples and Methods

3.1 Preface

This chapter presents a sequence of example networks, and develops techniques for applying the HHD both to the examples and to generic networks. First, a series of simple examples are presented (see Section 3.2). These examples are chosen either for their importance, or to illustrate the mechanics of the decomposition. Next, in Section 3.3, we develop methods for applying the HHD to the Cartesian products of graphs. Graphs that arise from Cartesian products are important in modeling applications, and by studying Cartesian products we develop a better understanding of lattices. Lattices are an important special case, which are addressed in Section 3.4. The chapter concludes by presenting general methods for applying the HHD to arbitrary networks (see Section 3.5).

3.2 Special Cases

In this section we consider a sequence of important special cases that are simple enough to solve explicitly. The cases are chosen to demonstrate the basics of the decomposition, and for their conceptual value.

3.2.1 Trees and Loops

A tree is a network with no closed loops. Trees are the simplest networks to analyze with potentials. Trees are automatically conservative since they contain no loops. Therefore $f = f_{\text{con}}$. An example tree is shown in Figure 3.1.

Since any edge flow on a tree is conservative the scalar potential ϕ can be recovered by setting the potential to zero at an arbitrary node, then summing f over the path from the initial node to every other node. If $\phi_1 = 0$ then ϕ_j is given by the work to move from node 1 to node j over the path from node 1 to node j in the tree. For example, the value of the potential at node 8 in Figure 3.1 equals $f_{12} + f_{25} + f_{58}$. Suppose we assign the edges reference orientations that point from lower indexed nodes to higher indexed nodes. Then $\phi_8 = f_A + f_D + f_G$. If the reference orientation on an edge is reversed, then the sign of f in the sum may have to be reversed. For example, if the convention on edge D is reversed so that it points from node 5 to 2, then $\phi_8 = f_A - f_D + f_G$.

Once the scalar potential has been found, we can always add a constant to ϕ. If, for example, we desire $\sum_j \phi_j = 0$ then should subtract the average value of the potential, $\frac{1}{V}\sum_j \phi_j$, from the potential.

The next simplest case to consider is a loop or cycle. The triangle is the simplest network which may have a nonzero scalar and vector potential, and is the smallest loop possible.

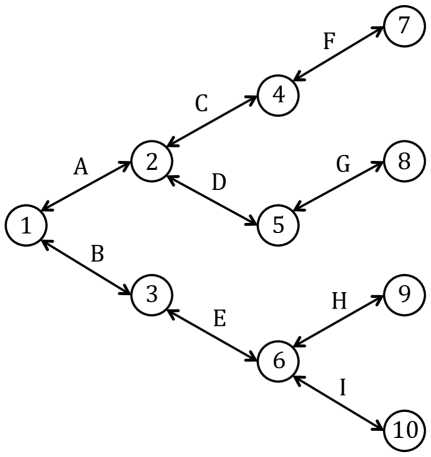

Figure 3.1: An example tree, with undirected edges.

Start by indexing the nodes from 1 to 3. Order and orient the edges 1 to 2, 2 to 3, 3 to 1. Index the faces of the network. In this case there are two, the inside of the triangle and the outside of the triangle. To ensure uniqueness we will assume the vector potential outside the triangle is zero, so will only consider the vector potential on the inner face. Choose positive rotation to correspond to the direction of the edges (1 to 2 to 3) as shown in Figure 3.2.

The corresponding Gradient is:

$$G = \begin{bmatrix} -1 & 1 & \cdot \\ \cdot & -1 & 1 \\ 1 & \cdot & -1 \end{bmatrix}.$$

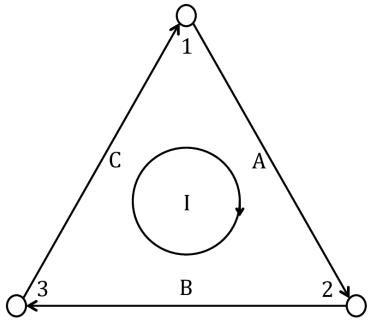

Figure 3.2: A triangle network. Arrows indicate the sign convention on each edge.

More generally, the gradient for a loop of V nodes is:

$$G = \begin{bmatrix} -1 & 1 & \cdot & \cdot \\ \cdot & -1 & 1 & \cdot \\ \cdot & \cdot & \ddots & \ddots \\ 1 & \cdot & \cdot & -1 \end{bmatrix}_{V \times V} .$$

The curl is trivial since there is only one cycle, and it crosses all edges in their positive direction:

$$C = [1, 1, \ldots, 1]$$

It follows that the curl transpose is $\mathbf{1}$, the vector of all ones. The curl transpose induces a simple mapping from θ to f_{rot}: $f_{\text{rot}} = \theta$ on every edge.

The corresponding node Laplacian (for a V state loop) is the standard second difference

operator for a line with periodic boundaries; the face Laplacian is a scalar:

$$L^2_{\mathcal{V}} = \begin{bmatrix} 2 & -1 & \cdot & -1 \\ -1 & 2 & -1 & \cdot \\ \cdot & \ddots & \ddots & \ddots \\ -1 & \cdot & -1 & 2 \end{bmatrix}_{V \times V}, \qquad L^2_{\mathcal{C}} = CC^\intercal = V.$$

The fact that the face Laplacian is a scalar makes the decomposition trivial. From the discrete Poisson equations $L^2_{\mathcal{C}}\theta = Cf$ so:

$$\theta = \frac{Cf}{V}.$$

Therefore, to compute the vector potential, compute the curl of f (sum f around the loop), then divide by the number of states. This may be thought of as an average of the total observed circulation. Then, by definition:

$$f_{\text{rot}} = C^\intercal\theta = \frac{C^\intercal Cf}{V}.$$

Note that $C^\intercal C$ is not the face Laplacian, $CC^\intercal = V$. Instead $C^\intercal C$ is an $V \times V$ matrix of all ones:

$$C^\intercal C = \mathbf{1}\mathbf{1}^\intercal = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}_{V \times V}.$$

By definition $f_{\text{con}} = f - f_{\text{rot}}$ so:

$$f_{\text{con}} = \left(I - \frac{1}{V} \mathbf{1}\mathbf{1}^{\intercal} \right) f.$$

That is, the conservative flow at edge k is f_k minus the average value of the edge flow over all V edges in the loop. This result provides a simple projection rule for finding the conservative field.

Given f_{con}, the scalar potential can be found trivially by solving:

$$-G\phi = f_{\text{con}}.$$

As usual, fix the potential at a node so that the corresponding linear system is triangular, and can be solved by back-substitution. This is equivalent to summing outward from an initial point on the loop.

Notice that this procedure is equivalent to the projection-integration method described in Section 2.3. First we take advantage of the simplicity of the rotational structure to compute θ, next we remove the rotational component, and compute the scalar potential by summing out from an initial node.

3.2.2 Singly Connected Components

A network has singly connected components if it is possible to break the network into separate components by removing an edge or a single node. A cut edge, or bridge, is an edge which, if removed, leaves two components separated. A cut node, articulation point, or separation point, is a node which, if removed[1], leaves two separate components [32, 33].

[1] along with all neighboring edges

A cut node can be transformed into a cut edge by removing it from the graph, then adding a copy of it to each connected component and adding an edge between the copies. A singly connected graph can be broken into biconnected components by removing all cut edges, or splitting all cut nodes. Here we assume that all cut nodes have been turned into cut edges, and the flow on any edge between copies of a cut node is set to zero.

Consider the set of cut edges. None of these edges are included in any basis cycle. If an edge is the only bridge between two components it is impossible to move from one component to the other and back without using that edge twice. It follows that f_{rot} is zero on all edges separating singly connected components.

Suppose edge k is a cut edge. Then, since f_{rot} is zero on all edges separating biconnected components, $f_{\text{con}k} = f_k$ so $\phi_{j(k)} - \phi_{i(k)} = f_k$. Now remove edge k from the graph, and apply the HHD to each of the separate components. The scalar potential on each component is only determined up to a constant, so we can always add a constant to the scalar potential on one of the components such that $\phi_{j(k)} - \phi_{i(k)} = f_k$.

It follows that, if a graph is singly connected, then it can be broken into its biconnected components, and the HHD can be applied to each component separately. Then the difference in potential across each edge separating biconnected components equals the flow over those edges, and the scalar potential for the full network can be recovered by adding a different constant to the scalar potential for each component chosen so that $\phi_{j(k)} - \phi_{i(k)} = f_k$ on edges k that are cut edges. Breaking a singly connected graph into its biconnected components can be done efficiently ($\mathcal{O}(V + E)$ time) [58, 59].

Thus, if all the loops in a network are edge disjoint (do not share any edges) then we can recover θ on each loop by applying the technique developed in the previous section for single loops. Evaluate the curl on the loop before dividing the curl by the perimeter of the loop (number of nodes in the loop). Then the rotational flow can be computed and

subtracted from the full edge flow to recover the conservative component, and the scalar potential can be computed by summing f out from an initial node over an arbitrarily chosen spanning tree.

If cut edges were formed by splitting cut nodes then, since there is no rotational flow over the cut edges, and the flow over the added cut edges is zero, the potential at a pair of nodes formed by splitting a cut node is equal. Thus the HHD on the original graph can be recovered by contracting the added edge, and setting the potential at the cut node to the potential at either side of the contracted edge.

3.2.3 Linked Loops: A Worked Example

So far we have only explicitly considered networks with single loops, or disjoint loops. In these cases it is trivial to compute the rotational potential and component, since θ is the curl on each loop divided by its perimeter. When the network includes multiple interacting loops the decomposition is more involved. The simplest example is a pair of triangles that share one edge. We consider this example in some detail to provide explicit examples of the operators, Laplacians, and projectors.

The node, edge, and face indexing is shown in Figure 3.3. The sign convention for each edge and cycle is shown by the direction of the arrows. The edge is indexing is chosen explicitly so that the outer loop is indexed consecutively.

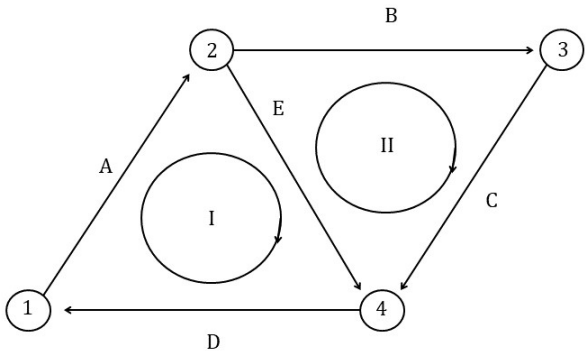

Figure 3.3: A pair of triangles sharing an edge. Arrows indicate the sign convention on each edge.

The corresponding gradient and node Laplacian are:

$$G = \begin{bmatrix} -1 & 1 & \cdot & \cdot \\ \cdot & -1 & 1 & \cdot \\ \cdot & \cdot & -1 & 1 \\ 1 & \cdot & \cdot & -1 \\ \cdot & -1 & \cdot & 1 \end{bmatrix}, \quad L^2_{\mathcal{V}} = \begin{bmatrix} 2 & -1 & \cdot & -1 \\ -1 & 3 & -1 & -1 \\ \cdot & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}.$$

Note that the node Laplacian is equal to diagonal degree matrix minus the adjacency matrix.

If we set the potential at the first node to zero the reduced node Laplacian is:

$$\hat{L}^2_{\mathcal{V}} = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}.$$

When constructing the adjoint curl we have the option of three possible cycle bases. We will use the cycle basis indicated in Figure 3.3. Then:

$$C^{\intercal} = \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \\ \cdot & 1 \\ 1 & -1 \\ 1 & \cdot \end{bmatrix}.$$

The face Laplacian can be computed by taking $CC^{\intercal}$, or since the network is planar, by evaluating the node Laplacian for the dual graph. The node Laplacain for the dual graph is given by computing the perimeter of each loop, and the shared perimeter between pairs of loops. Then the face Laplacian has negative off-diagonal entries equal to the shared perimeter of pairs of distinct loops, and positive diagonal entries equal to the perimeter of each basis loop:

$$L^2_{\mathcal{C}} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}.$$

Now suppose we are given an edge flow f. Since the system is small and meant as an example we might as well solve directly using the discrete Poisson equations:

$$\phi = -[L^2_{\mathcal{V}}]^{\dagger} G^{\intercal} f$$
$$\theta = [L^2_{\mathcal{C}}]^{-1} C f.$$

The inverse of the reduced node Laplacian and face Laplacian are:

$$[\hat{L}^2_{\mathcal{V}}]^{-1} = \frac{1}{8}\begin{bmatrix} 5 & 4 & 3 \\ 4 & 8 & 4 \\ 3 & 4 & 5 \end{bmatrix}, \quad [L^2_{\mathcal{M}}]^{-1} = \frac{1}{8}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

The columns of these matrices are analogous to Green's functions in the continuum.

Given G and C we can also compute the projectors onto f_{rot} and f_{con}. The projector onto the rotational space is:

$$P_{\text{rot}} = \frac{1}{8}\begin{bmatrix} 3 & 3 & 1 & 1 & 2 \\ 3 & 3 & 1 & 1 & -2 \\ 1 & 1 & 3 & 3 & -2 \\ 1 & 1 & 3 & 3 & 2 \\ 2 & -2 & -2 & 2 & 4 \end{bmatrix}.$$

This projector nicely reflects the structure of the network. The first pair of edges only neighbor loop I, and the second pair of edges only neighbor loop II, so f_{rot} must be constant on edges A and B, and edges C and D. It follows that the first two rows of the projector are identical, and the second two rows are identical. Only the last row is different since only the last row borders both loops.

The projector onto the space of conservative f can be computed from the first three columns of the QR factorization of G:

$$P_{\text{con}} = \frac{1}{8}\begin{bmatrix} 5 & -1 & -1 & -3 & -2 \\ -1 & 5 & -3 & -1 & 2 \\ -1 & -3 & 5 & -1 & 2 \\ -3 & -1 & -1 & 5 & -2 \\ -2 & 2 & 2 & -2 & 4 \end{bmatrix}.$$

As before, the fifth row and fifth column are distinct from the rest of the projector, since they correspond to the fifth (shared) edge which is distinct from the other four edges.

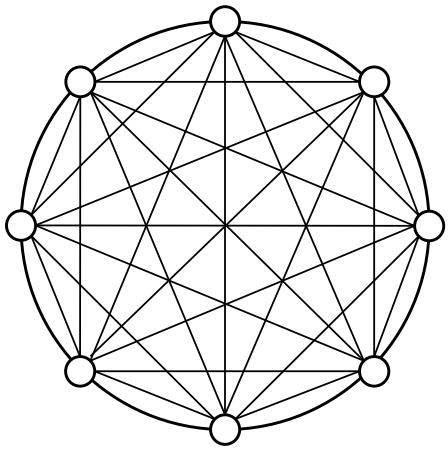

Figure 3.4: K_8, the complete graph on 8 vertices.

3.2.4 Complete Graphs

Consider a complete graph on V vertices. An example on 8 nodes is shown in Figure 3.4.

The corresponding node Laplacian has the simple form:

$$G^\intercal G = VI - \mathbf{1}\mathbf{1}^\intercal$$

where $\mathbf{1}$ is the vector of all ones.

This matrix has a very simple spectrum because adding xI to any matrix simply shifts its spectrum by x. The spectrum of $-\mathbf{1}\mathbf{1}^\intercal$ consists of one eigenvalue equal to $-V$ corresponding to $\mathbf{1}$, and $V-1$ eigenvectors equal to zero. The last $V-1$ eigenvalues are all zero since $-\mathbf{1}\mathbf{1}^\intercal$ is a rank one matrix. Any choice of $V-1$ orthogonal vectors that span the space perpendicular to $\mathbf{1}$ serve as eigenvectors. Denote this basis Q. Then $[\mathbf{1}, Q]^{-1} = [\mathbf{1}, Q]^\intercal$. Adding V to the spectrum means that the first eigenvalue is zero, and the rest are all equal to V. This means:

$$G^\intercal G = VQQ^\intercal.$$

Note that $QQ^\intercal$ is the orthogonal projector onto the space orthogonal to $\mathbf{1}$. By default $G^\intercal f$ has no projection onto $\mathbf{1}$ since $\mathbf{1}^\intercal G^\intercal f = f^\intercal G\mathbf{1}$ and $G\mathbf{1} = 0$. Therefore the discrete Poisson equation $-G^\intercal G\phi = G^\intercal f$ is solved by:

$$\phi = \frac{-1}{V} G^\intercal f = \frac{1}{V} Df. \tag{3.1}$$

Therefore the scalar potential at node i equals the average of f over all edges leaving node i:

$$\phi_i = \frac{\sum_{i \neq j} f_{ij}}{V}. \tag{3.2}$$

This conclusion will play an essential role in the study of competitive tournaments in Chapter 4 and Chapter 5.

It follows that the conservative flow equals $-\frac{1}{V} GG^\intercal f$ and the rotational flow equals $f - f_{\text{con}} = (I + \frac{1}{V} GG^\intercal) f$.

To solve for θ we need to pick a cycle basis. For a complete graph it is natural to start with the spanning tree associated with a single node (node 1). Then the set of all edges leaving node 1 form the tree, and the list of edges between $i, j \neq 1$ form the chords. There are exactly $(V-1)(V-2)/2$ of these chords. Let θ_{ij} denote the vector potential on the associated loops. This is a natural choice of basis since all the loops are triangles, and each chord is only included in one loop. It follows that θ_h equals $f_{\text{rot}\,k}$ on the edge k corresponding to chord h. Alternatively we could extend the cycle basis to include all triangles in the complete graph and solve the associated l_1 minimization problem for the sparsest representation on a triangle basis.

Therefore, for any complete graph the decomposition is remarkably simple. Rescale the divergence of f by $1/V$ to recover the scalar potential, then either s the rotational potential equal to the rotational flow on the chords left over after removing the conservative flow specified by the scalar potential, or use a linear programming solver to find the set of triangles which represents the rotating flow most concisely.

3.3 Graph Products

An important class of graphs in applications are graphs that are constructed from the products of multiple smaller graphs (cf. [60, 61, 62]). Graph products are a versatile set of binary operations on graphs [63]. Graph products arise naturally when studying graphs with repeated, or regular, structures [61]. There are three natural graph products which are widely used, the Cartesian product, the tensor product, and the strong product [64, 65]. This chapter focuses on Cartesian products. Here we analyze how the spaces and operators involved in the HHD of a product graph depend on the factor graphs, and introduce a spectral method for performing the HHD of a product graph based on the spectrum of the Laplacian on the factor graphs. This technique is especially useful for performing the decomposition on lattices, hypercubes, and products involving complete graphs.

3.3.1 Cartesian Products: Topology

This section introduces a systematic method for understanding the cycle space of the Cartesian products of graphs based on the cycle and edge spaces of the smaller graphs used in the product.

Consider the Cartesian product of two graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$. Let $\mathcal{V}_1$ be the set of vertices of the first graph, $\mathcal{E}_1$ be its edges, and $\mathcal{C}_1$ be a cycle basis. Then let V_1,E_1, and L_1 be the

number of nodes, edges, and loops in the cycle basis of the first graph. In the same way, let $\mathcal{V}_2$ be the set of vertices of the second graph, $\mathcal{E}_2$ be its edges, $\mathcal{C}_2$ be its cycle basis, and let V_2, E_2, and L_2 be the number of nodes, edges, and loops in the cycle basis of the second graph.

The Cartesian product of two graphs is an extension of the notion of the Cartesian product of two sets originally introduced by Sabidussi [66]. Given sets A and B the Cartesian product of A and B, denoted $A \times B$, is the set of all possible pairs of elements of A and elements of B. That is, if $a \in A$ and $b \in B$ then $A \times B$ is the set of all pairs of the form a, b. To motivate the use of "Cartesian" notice that, if the set A is all real numbers, and the set B is also all real numbers, then the Cartesian product of A and B is the set of all Cartesian coordinates of points in $\mathbb{R}^2$.

The Cartesian product of two graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$, is denoted:

$$\mathcal{G} = \mathcal{G}_1 \square \mathcal{G}_2. \tag{3.3}$$

Note that the Cartesian product is denoted with a square rather than with $\times$, which denotes the Cartesian product for sets. This is to reserve the use of $\times$ for the tensor product [64]. The product graph $\mathcal{G}$ has one vertex for each pair of vertices from the factor graphs:

$$\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2 = \{\text{all pairs } (v, w) : v \in \mathcal{V}_1, w \in \mathcal{V}_2\} \tag{3.4}$$

and edges between states $\{v_i, w_k\}$ and $\{v_j, w_h\}$ if there is an edge in $\mathcal{E}_1$ between v_i and v_j and $h = k$, or $i = j$ and there is an edge in $\mathcal{E}_2$ between w_k and w_h. That is, the Cartesian product of two graphs has state space equal to the Cartesian product of the state spaces of the two graphs, and edges between pairs of states that differ in only one of the two states in each pair, and the states that differ are connected by an edge in their corresponding graph

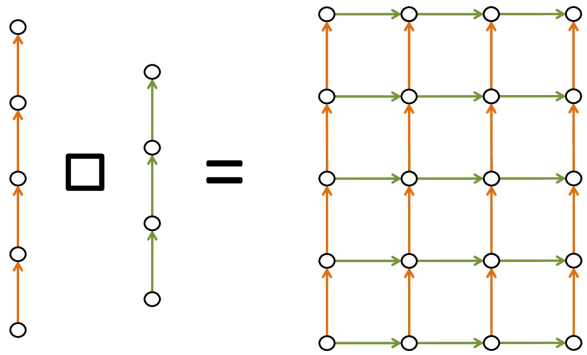

Figure 3.5: The Cartesian product of two lines produces a grid. The first line has 5 states and the second has 3 so the grid is 5 by 3. Notice that the edges in the product only change one coordinate at a time.

[61, 64].

This definition can be motivated by the following simple example. Consider a pair of interacting populations. Each population can take on a number of states, i.e. the number of individuals in the population. Then each of the two populations can take on states in $\mathbb{Z}$, and the full state space is $\mathbb{Z}^2 = \mathbb{Z}\square\mathbb{Z}$. Suppose that, in addition, we only allow for edges corresponding to single birth and death events. This requires that only one population can change states at a time. This is not a restrictive requirement since, in continuous time, the probability any two birth or death events occur simultaneously is zero. Then, the set of edges of the full graph are edges corresponding to a birth in the first population with the second population fixed, a death in the first population with the second population fixed, a birth in the second population with the first population fixed, or a death in the second population with the first population fixed. This is exactly the set of edges given by the Cartesian product. Note that this does not require that the process on each population is independent of the others since the rates of birth and death in a given population may depend on the number of individuals in the other population, as when two populations compete for the same resources, or one population consumes the other. Graphically this

maps $\mathbb{Z}\square\mathbb{Z}$ onto the lattice $\mathbb{Z}^2$ with strictly horizontal and vertical edges, each of length 1, as shown in Figure 3.5.

The Cartesian product is both commutative, and associative [66]:

$$\begin{aligned} \mathcal{G}_1\square\mathcal{G}_2 &= \mathcal{G}_2\square\mathcal{G}_1 \\ \mathcal{G}_1\square\mathcal{G}_2\square\mathcal{G}_3 &= (\mathcal{G}_1\square\mathcal{G}_2)\,\square\mathcal{G}_3 = \mathcal{G}_1\square\,(\mathcal{G}_2\square\mathcal{G}_3) \end{aligned} \tag{3.5}$$

so there is no ambiguity in taking the Cartesian product of more than two graphs. The product of more than two graphs can be broken into a sequence of pairs of products, and the result does not depend on the order of that sequence. For example, taking the Cartesian product of $\mathbb{Z}^2$ with $\mathbb{Z}$ would produce $\mathbb{Z}^3$, which is also the Cartesian product of $\mathbb{Z}$ with itself three times. In this way all d dimensional lattices can be viewed as the Cartesian product of $\mathbb{Z}$ with itself d times.

The Cartesian product is analogous to multiplication because it is equivalent to starting with $\mathcal{G}_1$, then, replacing every state in $\mathcal{G}_1$ with a distinct copy of $\mathcal{G}_2$. For example, the Cartesian product of a triangle with a triangle can be constructed by starting with a triangle, then introducing a triangle for every node in the previous triangle. A familiar example is the construction of a hypercube from the Cartesian product of a cube with a line. These two examples are illustrated in Figure 3.6.

Like multiplication on the integers there are also graphs corresponding to one and zero, the graph with one node and the graph with no nodes, and a limited notion of divisibility. If $G = G_1\square G_2$ then G divided by G_1 is G_2. Like the integers there are prime graphs that cannot be expressed as the Cartesian products of any two graphs [66]. More strongly, any connected graph that is a Cartesian product of graphs can be uniquely factorized into prime graphs [64, 66]. A method for factoring product graphs is presented in [67].

The goal of this section is to present a method for constructing the cycle space of a

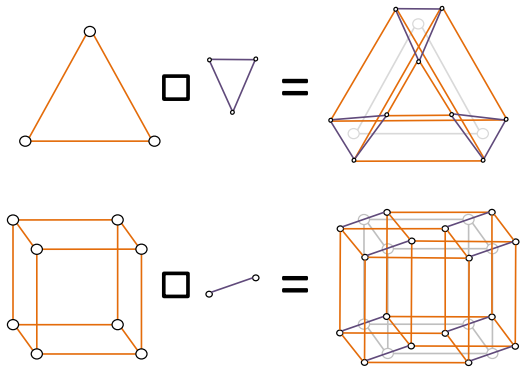

Figure 3.6: The Cartesian product of two triangles, and the Cartesian product of a cube with a line. In both cases the Cartesian product is constructed by replacing each node in the orange graph with the purple graph.

Cartesian product of graphs based only on the cycle spaces of the factor graphs, and pairs of edges drawn from pairs of factor graphs. In Section 3.3.2 we present an algorithm for constructing the gradient, node Laplacian, curl, and face Laplacian for the Cartesian product of graphs. This will be an essential tool for studying large graphs that can be expressed as products of small factor graphs. Then the operators for the large graph can be constructed automatically from the operators for the smaller graph. It will also present a more ordered understanding of how the operators for lattices, and will enable an elegant spectral approach to the decomposition (see Section 3.3.3).

To start, consider the dimension of the node, edge, and loop spaces of the product graph $\mathcal{G} = \mathcal{G}_1 \square \mathcal{G}_2$. After the product there are exactly $V = V_1 V_2$ states in the graph $\mathcal{G}$. This follows immediately from the definition of the Cartesian product on sets.

To find the number of edges note that every edge in $\mathcal{E}_1$ can be uniquely identified by its endpoints v_1, v_2. Then, after the Cartesian product this particular edge is copied V_2 times for each possible set of pair $(v_1, w), (v_2, w)$ where $w \in \mathcal{V}_2$. The same argument extends to all edge in $\mathcal{E}_2$, only each is copied V_1 times. Therefore the total number of edges in the

product is:

$$E = E_1 V_2 + E_2 V_1. \tag{3.6}$$

Given the number of edges and nodes in the product we can immediately compute the dimension of the cycle space using the standard formula for the cyclomatic number:

$$L = E - V + 1 = E_1 V_2 + E_2 V_1 - V_1 V_2 + 1. \tag{3.7}$$

The cyclomatic number can be written more suggestively by replacing E_1 and E_2 with $L_1 + V_1 - 1$ and $L_2 + V_2 - 1$. Then:

$$\begin{aligned} L &= (L_1 + V_1 - 1)V_2 + (L_2 + V_2 - 1)V_1 - V_1 V_2 + 1 \\ &= [L_1 V_2 + L_2 V_1] + [(V_1 - 1)(V_2 - 1)]\,. \end{aligned} \tag{3.8}$$

Written in this form the cycle space dimension, L, suggests a decomposition into two classes of loops. The first are formed by the sets circuits in the product graph formed by holding one component fixed, and moving about a loop in the cycle basis for the other component. For example, if the loop $1 \to 2 \to 3 \to 1$ is a loop in $\mathcal{G}_1$ then there are V_2 corresponding loops in $\mathcal{G}$ corresponding to loops of the form $(1, w) \to (2, w) \to (3, w) \to (1, w)$ where $w \in \mathcal{V}_2$. There are exactly $L_1 V_2$ loops of this kind. They are clearly independent of each other since the set of all loops in this set with the second component fixed is the set of basis loops of $\mathcal{G}_1$, and the V_2 copies of these sets are all disjoint. These sets are disjoint since they only use edges that change the first component, so cannot connect any states who differ in their second component. As an example of a set of loops of this type look at the three purple loops that appear in the Cartesian product

of an orange and purple triangle shown in Figure 3.6. By the same logic we can produce a set of F_2V_1 linearly independent loops by copying the cycle basis of $\mathcal{G}_2$ V_1 times. These two sets of loops are also disjoint since the first only uses edges between states that differ in their first component, and the second only uses edges between states that differ in their second component.

Therefore, the set of loops that are given by fixing one component, and then carrying the other around a basis loops in $\mathcal{C}_1$ or $\mathcal{C}_2$ is a set of $L_1V_2+L_2V_1$ linearly independent loops. Denote this set $\mathcal{C}_{\text{factor}}$ since it is the direct extension of the cycle basis of each factor graph. Formally:

$$\mathcal{C}_{\text{factor}} = (\mathcal{C}_1 \times \mathcal{V}_2) \cup (\mathcal{V}_1 \times \mathcal{C}_2) \tag{3.9}$$

where $\times$ denotes the standard Cartesian products of sets.

The set $\mathcal{C}_{\text{factor}}$ accounts for the first $L_1V_2 + L_2V_1$ loops in $\mathcal{C}$, but does not account for the remaining $(V_1-1)(V_2-1)$ loops. To see why there are more loops in $\mathcal{C}$ than are spanned by combinations of the circuits in $\mathcal{C}_{\text{factor}}$, glance back at the Cartesian product of a line segment with another line segment shown in Figure 3.5. Both factor graphs have no loops, yet the Cartesian product of the two graphs is full of cycles. These cycles are formed by picking on edge from $\mathcal{G}_1$ and another from $\mathcal{G}_2$. In general the cycles that are formed in this way take a characteristic form: a square. This is the motivation for the use of the square as the symbol for the operator: the Cartesian product of two edges is a square. The space of loops spanned by the Cartesian product of pairs of edges is the square space of the graph [64].

Loops of this kind are always formed by picking one edge from $\mathcal{E}_1$ and one edge from $\mathcal{E}_2$. Denote the endpoints of the first edge v_1, v_2 and the second w_1, w_2. Then the corresponding loop is constructed by walking across the first edge, across the second edge, backwards across the first, then backwards across the second. This moves from (v_1, w_1) to

(v_2, w_1) to (v_2, w_2) to (v_1, w_2) and finally back to (v_1, w_1).

Clearly there are E_1E_2 squares in the square space, however we are only looking for $(V_1 - 1)(V_2 - 1)$ loops to complete the cycle basis. Since, in general $E_1 \geq V_1 - 1$ and $E_2 \geq V_2 - 1$ the set of all loops formed by the Cartesian product of two edges is, usually too large. In fact, unless both factor graphs are trees this set is too large. The special case when both factor graphs are trees offers inspiration for a method for finding a cycle basis for the square space.

It is always possible to construct a cycle basis of a graph by first picking a spanning tree for the graph, and then associating each cycle with a particular chord. Assume that $\mathcal{C}_1$ and $\mathcal{C}_2$ are fundamental cycle bases associated with spanning trees $\mathcal{T}_1$ and $\mathcal{T}_2$. The spanning trees have $V_1 - 1$ and $V_2 - 1$ edges respectively. Therefore, if we prune $\mathcal{G}_1$ and $\mathcal{G}_2$ down to a pair of spanning trees $\mathcal{T}_1$, $\mathcal{T}_2$ associated with their loop bases then we have identified two sets of edges of size $V_1 - 1$ and $V_2 - 1$. Therefore the Cartesian product of these two trees will have a loop space with dimension $(V_1-1)(V_2-1)$, which is exactly the right size to complement the loop space formed by the independent loop spaces of the factor graphs. Denote the loops space of the Cartesian product of these two trees $\mathcal{C}_{\text{tree}}$.

It remains to show that the cycle space of the Cartesian product can be decomposed into a subspace of circuits spanned by circuits from $\mathcal{C}_{\text{factor}}$, and circuits from $\mathcal{C}_{\text{tree}}$. That is, to show that:

$$\mathcal{C} = \mathcal{C}_{\text{factor}} \cup \mathcal{C}_{\text{tree}} \tag{3.10}$$

is a cycle basis.

By construction $\mathcal{C}_{\text{tree}}$ consists of $(V_1 - 1)(V_2 - 1)$ loops. This set is independent from the set of loops $\mathcal{C}_{\text{factor}}$ since all the loops in $\mathcal{C}_{\text{factor}}$ must include a chord of either $\mathcal{G}_1$ or $\mathcal{G}_2$, and none of the loops in $\mathcal{C}_{\text{tree}}$ include a chord since the chords were pruned to produce the

trees. This implies that every loop in $\mathcal{C}_{\text{factor}}$ includes at least one edge that is not included in any loop in $\mathcal{C}_{\text{tree}}$. Therefore there is no way to add the loops in $\mathcal{C}_{\text{tree}}$ together to produce a loop in $\mathcal{C}_{\text{factor}}$. Alternatively, every linear combination of loops in $\mathcal{C}_{\text{factor}}$ must include a chord, so cannot produce any loop in $\mathcal{C}_{\text{tree}}$.

All that remains to show is that $\mathcal{C}_{\text{tree}}$ is a linearly independent set of cycles. This independence can be shown as follows. First, index all the nodes in both trees, moving outward from some root. Next index all the edges in each tree in lexicographical order of their endpoints. Then list all the pairs of edges that form the loops in $\mathcal{C}_{\text{tree}}$ in lexicographical order of the indexes corresponding to the two edges that define each loop. This defines an ordering of all the loops in $\mathcal{C}_{\text{tree}}$ such that the $h + 1^{st}$ loop in the set must always contain a node $(v, w) \in \mathcal{V}$ that was not contained in any of the previous loops. Since all loops must have an edge into, and out of every node they pass through, this implies that the $h + 1^{st}$ loop contains a pair of edges that is not contained in any of the previous loops. Therefore the $h + 1^{st}$ loop is independent of all the previous loops. In turn this implies that $\mathcal{C}_{\text{tree}}$ is a linearly independent set of loops, so has dimension $(V_1 - 1)(V_2 - 1)$, and the loop space $\mathcal{C}$ can be decomposed into the basis formed by the loops in $\mathcal{C}_{\text{factor}}$ and $\mathcal{C}_{\text{tree}}$.

Note that the cycle bases on the components need not be a fundamental cycle bases since any cycle basis can be reached by a linear combination of cycles in the fundamental cycle bases associated with the trees used to build $\mathcal{C}_{\text{tree}}$ without changing the range of the cycles in $\mathcal{C}_{\text{factor}}$

This provides a general construction rule for building the cycle space of Cartesian products of two graphs. First, construct a spanning tree for both networks, and a cycle basis for both networks. Then consider all loops that are formed by Cartesian products of an edge in the first tree and an edge in the second. Next consider all loops that are given by fixing a component, and picking a loop from the cycle space of the remaining factor

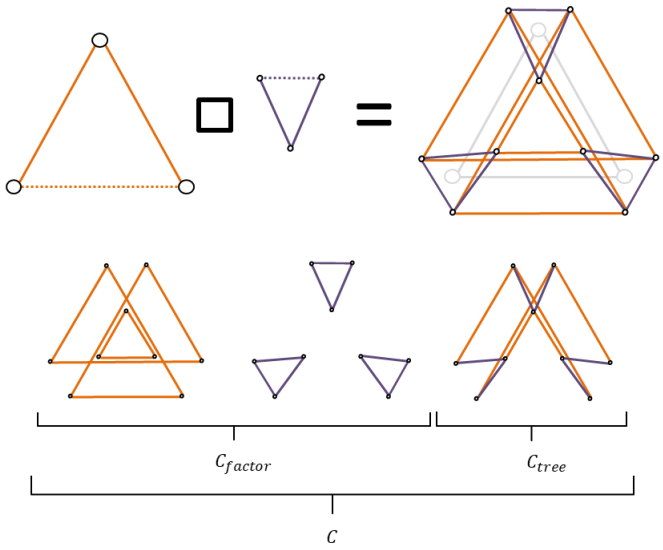

Figure 3.7: The cycle basis for the Cartesian product of two triangles. The chords that are pruned to produce the spanning trees are shown with dashed lines.

graph. This particular basis is the Hammack basis [44, 64] of a product graph. It should be noted that the Hammack basis usually isn't the smallest cycle basis in the sense that there are usually other bases with a smaller total perimeter [44]. However, for our purposes this construction is preferred because it introduces a natural partition of the loop space into two sets, and leads to an easy construction rule for the curl and face Laplacian.

For example, consider the cycle space associated with the Cartesian product of two loops. The resulting graph is a square lattice on a torus. Then the set of loops in $\mathcal{C}_{\text{factor}}$ are the sets of circles that wrap around the torus. A grid on a torus can always be represented as a grid with periodic boundary conditions. The spanning trees for both loops are lines, so the Cartesian product of the two spanning trees is always a grid. The chord to be removed can be chosen to be the edge that passes around the periodic boundary of the grid when the torus is represented as a grid with periodic boundaries. Then $\mathcal{C}_{\text{tree}}$ is just the set of faces of the grid when the edges crossing the periodic boundary are removed. This division of the

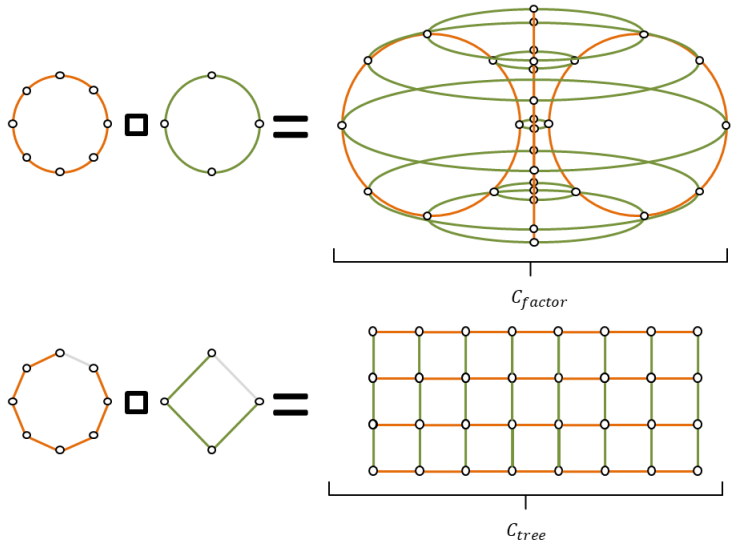

Figure 3.8: The Cartesian product of two loops is a grid on a torus. The component loops correspond to the loops that circumnavigate the torus (shown in green and orange on the torus), and the loops formed by the Cartesian product of the two spanning trees are the faces of the grid that is formed when the torus is cut and laid flat as a grid. The chords of the spanning tree correspond to the edges that are cut.

loops on a torus is shown in Figure 3.8.

The Hammack basis can be used to build a basis for the cycle space of the Cartesian products of more than two graphs by breaking the product into a sequence of sequential products. For example, given $\mathcal{G} = \mathcal{A}\square\mathcal{B}\square\mathcal{C}$ we could construct $\mathcal{C}$ by first constructing a basis for $\mathcal{A}\square\mathcal{B}$, then applying the method again to find a basis for $(\mathcal{A}\square\mathcal{B})\square\mathcal{C}$. That is, first find a basis for the cycle space of $\mathcal{A}$ and $\mathcal{B}$, then construct the set of loops $\mathcal{C}_{\text{factor}}(A\square B)$ formed by picking a basis loop from $\mathcal{C}_A$ or $\mathcal{C}_B$, and a state from $\mathcal{V}_B$ or $\mathcal{V}_A$ respectively. Then pick a spanning tree for $\mathcal{A}$ and $\mathcal{B}$, and construct the set of all loops formed by taking the Cartesian product of an edge drawn from the spanning tree on $\mathcal{A}$ and the spanning tree on $\mathcal{B}$. This forms a basis for the cycle space of the Cartesian product of $\mathcal{A}$ and $\mathcal{B}$. To construct the full basis we use this Hammack basis as the cycle basis for $\mathcal{A}\square\mathcal{B}$ that is required to build $\mathcal{C}_{\text{factor}}((A\square B)\square C)$. To complete the Hammack basis of $\mathcal{A}\square\mathcal{B}\square\mathcal{C}$ we need a spanning

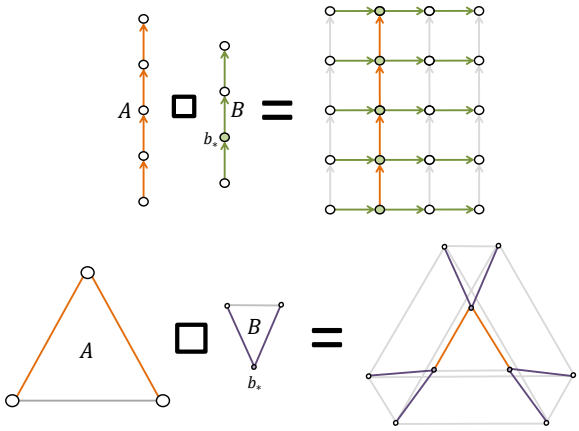

Figure 3.9: The spanning trees for two examples of Cartesian graph products. Edges not in the trees are shown in grey. The special fixed node in $\mathcal{B}$ is denoted b_*, and the corresponding nodes are shaded in.

tree for $\mathcal{A}\square\mathcal{B}$.

This tree can be constructed directly from the spanning trees for $\mathcal{A}$ and $\mathcal{B}$. First, fix the second component of the pair of states (a, b) at a particular node b_* in $\mathcal{B}$. Then we can span all different states of the form (a, b_*) with the spanning tree for $\mathcal{A}$. This allows us to move from any state (a_1, b_*) to (a_2, b_*). What remains is a tree that connects any other point in the product space. Suppose we want to move from (a_1, b_1) to (a_2, b_2). Then hold a_1 fixed, and walk along the spanning tree of $\mathcal{B}$ to get to (a_1, b_*). Then hold b_* fixed and walk along the spanning tree of $\mathcal{A}$ to get to (a_2, b_*). Then, holding a_2 fixed, walk along the spanning tree of $\mathcal{B}$ to get to (a_2, b_2). An example is shown in Figure Figure 3.9 for the Cartesian product of two line segments, and the Cartesian product of two triangles. Notice that this construction is not commutative, and requires the choice of some special node in $\mathcal{B}$.

Given a spanning tree for the Cartesian product of $\mathcal{A}$ and $\mathcal{B}$ we can build $\mathcal{C}_{tree}$ for $(\mathcal{A}\square\mathcal{B})\square\mathcal{C}$, and a Hammack basis for $\mathcal{G}$. Notice that this construction depends on the order in which the products are applied.

Repeated Products and Lattices

The technique developed for constructing Hammack bases of product graphs is especially useful when considering the repeated Cartesian product of a graph with itself. Consider the product:

$$\mathcal{G} = \mathcal{G}_0^n = \mathcal{G}_0^{n-1} \square \mathcal{G}_0 = \mathcal{G}_0 \square \mathcal{G}_0 ... \square \mathcal{G}_0 \tag{3.11}$$

A product of this kind is the generalized hypercube of $\mathcal{G}_0$ [64]. If $\mathcal{G}_0$ then the graph is a Hamming graph [68]. The lattice $\mathbb{Z}^n$ is the generalized hypercube of the line.

Before specifying the construction of the cycle space it is important to consider the asymptotics of the dimension of each space. Let V_0, E_0, F_0 denote the dimensions of the node, edge, and loop space in the original graph. Then, after n products with itself:

$$\begin{aligned} V &= V_0^n \\ E &= nE_0V_0^{n-1}. \end{aligned} \tag{3.12}$$

The number of edges can be computed by noting that any edge in the product can be specified by first picking which copy of $\mathcal{G}_0$ it is drawn from (n choices), then picking an edge from $\mathcal{G}_0$ (E_0 options), then specifying the state of the remaining $n-1$ compartments (V_0^{n-1} options). It follows that the loop space has dimension:

$$L = (nE_0 - V_0)V_0^{n-1} + 1 = nL_0V_0^{n-1} + (n-1)V_0^n - nV_0^{n-1} + 1. \tag{3.13}$$

The range of gradient spans $(V-1)/E$ percent of the space of edges, so, as n becomes large the percent of the edge space spanned by the gradient vanishes $\mathcal{O}(1/n)$:

$$\frac{V}{E} = \frac{V_0^n - 1}{nE_0V_0^{n-1}} \to \frac{V_0}{nE_0} \to 0. \tag{3.14}$$

Consequently, as n becomes large the loop space grows much faster than the node space. The range of adjoint curl spans L/E percent of the space of edges. Since $L = E - (V - 1)$, $\frac{F}{E}$ converges to $1 - \frac{V_0}{nE_0}$ for large n. Therefore, when n is large the loop space makes up all but $\mathcal{O}(1/n)$ percent of the edge space. Therefore, even if the dimension of the initial cycle space is small, the dimension of the cycle space grows faster than the dimension of the state space in n, and for $n > \frac{V_0}{E_0}$ the dimension of the cycle space will be larger than the number of vertices in the network.

It is also interesting to consider the proportion of the loop space associated with $\mathcal{C}_{\text{factor}}$ and with $\mathcal{C}_{\text{tree}}$. When n is large:

$$\frac{|\mathcal{C}_{\text{factor}}|}{L} = \frac{nL_0V_0^{n-1}}{(nE_0 - V_0)V_0^{n-1} + 1} \to \frac{L_0}{E_0} + \mathcal{O}(1/n). \tag{3.15}$$

Therefore, as n becomes large, the fraction of loops in the cycle basis that are copies of loops in the original graph approaches F_0/E_0 from above. Therefore, if E_0 is much larger than V_0 (the initial graph is dense) then most of the loops in the Hammack basis of $\mathcal{G}$ will be loops from the factor graphs, and if $E_0 \approx V_0$ then most of the loops will be squares in the square space of the product graph.

For concreteness we will now consider an important example: the Cartesian product of a line segment with itself. The example will help show how the construction developed at the start of this chapter is an essential tool for understanding the cycle basis of high dimensional lattices.

Set $\mathcal{G}_0$ to a line segment with $V_0 = m$ nodes. Then $E_0 = m - 1$ and $F_0 = 0$. The first product $\mathcal{G}_0^2$ is simply the m by m grid. Suppose $m = 3$. Then $\mathcal{G}_0^2$ has 9 nodes, 12 edges, and 4 basis loops (see Figure 3.10). Notice that in this simple example the number of basis loops is exactly equal to the number of faces in the graph, which is exactly the number of

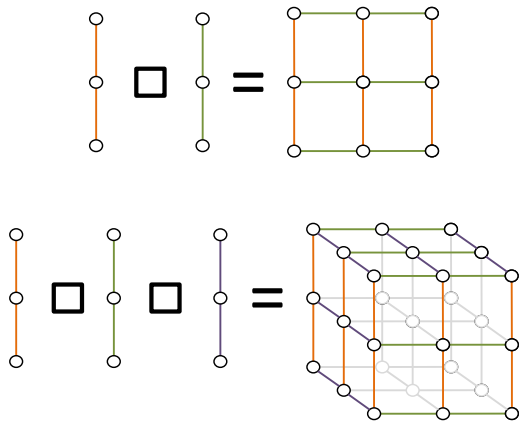

Figure 3.10: The Cartesian product of a line segment with itself producing a 2D and 3D lattice.

pairs of edges that can be chosen from $\mathcal{G}_0$. This is because $\mathcal{G}_0$ has no loops, so the spanning tree for $\mathcal{G}_0$ is $\mathcal{G}_0$. From this example one might conjecture that the number of basis loops in a lattice is equal to the number of squares in the lattice, and the basis can be constructed directly from these squares. This is true for the 2D lattice, but is not true for the 3D lattice, or any higher dimensional lattice. For example, the three dimensional lattice, $\mathcal{G}_0^3$, has 27 nodes, and 54 edges, leaving 28 basis loops. The lattice has 36 squares, not 28, so the set squares is too large to be a cycle basis.

The fact that the space of basis loops does not include all squares follows from the fact that the cycle basis of a product graph does not include all loops formed by pairs of edges in the factor graphs. The 3D lattice can be written $\mathcal{G}_0^2 \square \mathcal{G}_0$ and $\mathcal{G}_0^2$ has a nonempty loop space. Therefore a spanning tree of $\mathcal{G}_0^2$ does not include all of its edges, so the space of basis loops of $\mathcal{G}_0^3$ does not include all loops that can be formed by pairs of edges.

A simpler example is the cube formed by taking the Cartesian product of an edge with itself three times. The cube has six faces, but only five independent loops. Any face of

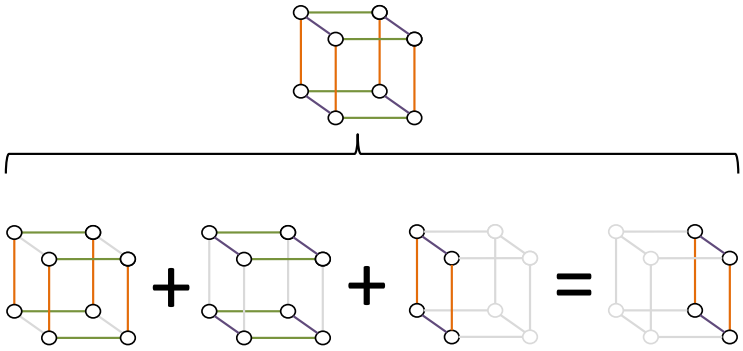

Figure 3.11: The Cartesian product of an edge with itself three times produces a cube. There are 6 faces of the cube, but the dimension of the loop space is five. Any face of the cube can be produced by adding together the remaining five faces.

the cube can be constructed by combining the other five faces of the cube, as illustrated in Figure 3.11. A cycle basis for the cube can be constructed by omitting one face from the set of faces. For example, pick all the vertically oriented faces (up and down), all the azimuthally oriented faces (front and back), and one horizontally oriented face (either left or right). This construction is a Hammack basis. The cube, $\mathcal{G}_0^3$, is the product of $\mathcal{G}_0^2$ and $\mathcal{G}$. Orient $\mathcal{G}_0^2$ so that it corresponds to either the top or bottom face, and $\mathcal{G}_0$ so that it corresponds to a vertical edge. Then $\mathcal{C}_{\text{factor}}$ are the top and bottom faces, and $\mathcal{C}_{\text{tree}}$ correspond to three of the four remaining faces. Which face is left out corresponds to which edge is left out of the spanning tree of $\mathcal{G}_0^2$. If the rightmost edge is left out then the set of basis loops is all the faces except the right face of the cube. The same type of construction works for the three by three by three lattice presented before (see Figure 3.12).

In effect, this construction partitions the squares in the lattice according to the orientation of the edges on the perimeter of the loop. Any edge only changes one coordinate at a

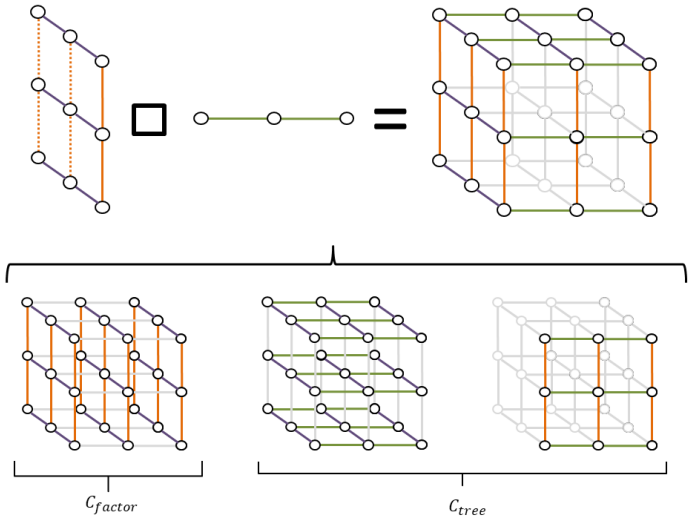

Figure 3.12: The cycle basis of a three by three by three lattice. The three by three by three lattice is shown in the upper right, with the edges left out of the spanning tree marked with dashes. The three by three by three lattice is formed by taking a product of the three by three lattice with a three node line segment. This lattice is shown in the upper right, with internal edges and nodes greyed out for ease of viewing. The bottom shows the cycle space, which can be broken into three sets of cross-sections. The first two include all azimuthal, and all vertical faces. The final is a sole cross-section given by fixing the azimuthal coordinate. The missing cross-sections correspond to the edges left out of the three by three lattice in its spanning tree.

time, so for an n dimensional lattice each possible face orientation is specified by a choice of a pair of coordinates. Let the (i, j, s) cross-section (or page) be the set of all loops given by picking a pair of edges, one which changes coordinate i and one which changes coordinate j, and then fixing the remaining coordinates to be equal $s_1, s_2, ...s_{n-2}$. Let the (i, j) book be the set of all cross-sections (pages) oriented in (i, j). Denote the (i, j) book $\mathcal{C}_{i,j}$ and the (i, j, s) cross section $\mathcal{C}_{i,j}(s)$. Then $\mathcal{C}_{i,j} = \cup_{s\in\mathcal{V}_0^{n-2}}\mathcal{C}_{i,j}(s)$. In this notation the cycle basis of the three by three by three lattice is $\{\mathcal{C}_{1,2}, \mathcal{C}_{1,3}, \mathcal{C}_{2,3}(1)\}$.

This cycle basis construction extends naturally to higher dimensional lattices. Suppose the lattice is n-dimensional. Then consider all loops contained in the library $\mathcal{C}_{1,2}$, $\mathcal{C}_{1,3}$, $\mathcal{C}_{1,4}$, ... $\mathcal{C}_{1,n}$. Each book in the library contains $E_0^2V_0^{n-2}$ loops. There are $n-1$ books in the

library, so the library contains $(n-1)E_0^2V_0^{n-2}$ loops. The full lattice had $(nE_0-V_0)V_0^{n-1}+1$ loops, however, since $\mathcal{G}_0$ is a tree, $E_0 = V_0-1$ so the full lattice has $(n(V_0-1)-V_0)V_0^{n-1}+1 = (n-1)V_0^n - nV_0^{n-1}+1$ loops. Subtracting the $(n-1)(V_0-1)^2V_0^{n-2}$ loops contained in the first library leaves $(n-2)V_0^{n-1} - (n-1)V_0^{n-2}+1$ loops unaccounted for. This is precisely the number of basis loops in $\mathcal{G}_0^{n-1}$. There are V_0 copies of $\mathcal{G}_0^{n-1}$ in $\mathcal{G}_0^n$, one for each possible value of the first coordinate. So, to specify the remaining loops, fix the first coordinate to a specific value (usually to 1). The choice of the fixed value corresponds to the choice of fixed value used in the construction of spanning trees of Cartesian products. Then the remaining graph is $\mathcal{G}_0^{n-1}$. This process can be repeated recursively until the remaining lattice is a one-dimensional line segment, which contains no loops.

Therefore the loop space of an n-dimensional square lattice can be partitioned into a sequence of libraries of the form:

$$\mathcal{C} = \begin{bmatrix} \mathcal{C}_{1,2}, \mathcal{C}_{1,3}, \mathcal{C}_{1,4}...\mathcal{C}_{1,n} \\ \mathcal{C}_{2,3}(s_1), \mathcal{C}_{2,4}(s_1), ...\mathcal{C}_{2,n}(s_1) \\ \mathcal{C}_{3,4}(s_1, s_2), ...\mathcal{C}_{3,n}(s_1, s_2) \\ \vdots \\ \mathcal{C}_{n-1,n}(s_1, s_2, ...s_{n-2}) \end{bmatrix}. \tag{3.16}$$

It remains to check that this set contains the right number of loops. Summing over each

library:

$$\begin{aligned}\sum_{j=1}^{n-1}(n-j)(V_0-1)^2V_0^{n-1-j} &= \frac{(V_0-1)^2}{V_0}\sum_{j=1}^{n-1}(n-j)V_0^{n-j} \\ &= \frac{(V_0-1)^2}{V_0}\frac{(n-1)V_0^{n+1}-nV_0^n+V_0}{(V_0-1)^2} \\ &= (n-1)V_0^n - nV_0^{n-1}+1\end{aligned}$$

which is precisely the dimension of the cycle space of the lattice.

Thus the Hammack cycle basis for pairwise products can be extended to provide an elegant decomposition of the loop space of high dimensional lattices. This decomposition proceeds recursively, first considering all loops that change the first coordinate, then fixing the first coordinate to a particular value and considering all loops that change the second coordinate with the first coordinate fixed. Proceed one coordinate at a time, with all previous coordinates fixed, and including all loops that vary the coordinate of interest. Once all loops varying that coordinate have been counted fix it to a set value. Continue until a set of basis loops has been specified.

3.3.2 Cartesian Products: Operators

In order to apply the HHD to a Cartesian product of two graphs we need a method for constructing the gradient, node Laplacian, curl, and face Laplacian of the product graph. The basic tool for the construction of the operators is the Kronecker product.

Kronecker Products and Sums: a Review

The Kronecker product of the matrices $A \in \mathbb{R}^{n_1,m_1}$ and $B \in \mathbb{R}^{n_2,m_2}$ is the block matrix [61, 69]:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & ... & a_{1m_1}B \\ a_{21}B & a_{22}B & ... & a_{2m_2}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_1,1}B & a_{n_1,2}B & ... & a_{n_1,m_1}B \end{bmatrix}. \tag{3.17}$$

The Kronecker product is associative and bilinear, but is not commutative. In general $A \otimes B = S^{\intercal}_{n_1,n_2}(B \otimes A)S_{m_1,m_2}$ where S_{n_1,n_2} and S_{m_1,m_2} are perfect shuffle matrices. A perfect shuffle matrix $S_{n,m}$ is the permutation matrix which exchanges the ordering:

$$(1,1),(1,2),...(1,n),(2,1),(2,2),...(2,n),...,(m,1),(m,2),...(m,n)$$

with the ordering:

$$(1,1),(2,1),...(m,1),(1,2),(2,2),...(m,2),...,(1,n),(2,n),...(m,n).$$

That is, $S_{n,m}$ is the permutation that would exchange a column-wise vectorization of a matrix with a row-wise vectorization of the matrix [69].

The Kronecker product also obeys the mixed product rule, which states that if A,B,C,D are all matrices, with dimensions such that AC and BD can be formed then [61]:

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \tag{3.18}$$

A number of other properties of the Kronecker product will be useful here. In particular, the spectrum is the product of the spectra (counted with multiplicity), and the singular

values are the products of the singular values. As a consequence the inverse and psuedo-inverse of the product is the product of the inverses and psuedo-inverses, the trace of the product is the product of the traces, and the determinant of the product of $A \in \mathbb{R}^{n\times n}$, and $B \in \mathbb{R}^{m\times m}$ is $\det(A)^m \det(B)^m$ [69]. Similarly, transpose of a Kronecker product is the product of the transposes.

The Kronecker sum is defined:

$$A \oplus B = A \times I_m + I_n \times B \tag{3.19}$$

where I_m and I_n are $m \times m$ and $n \times n$ identity matrices and A is $n \times n$ and B is $m \times m$. The Kronecker sum has the convenient property that:

$$\exp(A \oplus B) = \exp(A) \otimes \exp(B). \tag{3.20}$$

We will show that the gradient, node Laplacain, and parts of the curl and face Laplacian associated with $\mathcal{C}_{\text{factor}}$ can all be formed by Kronecker products and Kronecker sums.

The Gradient and Node Laplacian

Consider the gradient first. The gradient maps from nodes to edges, so to specify the gradient we need to first specify an ordering for the nodes and for the edges. List the nodes lexicographically. This orders the nodes of the product:

$$(1,1),(1,2),...,(1,V_2),(2,1),(2,2),...(2,V_2),...(V_1,1),(V_1,2),...(V_1,V_2).$$

Notice that in this ordering all the edges from $\mathcal{G}_2$ exclusively connect nodes of the product graph, $\mathcal{G}$, whose indices are the same modulo V_2, and no edge from $\mathcal{G}_1$ connects nodes of

the product graph seperated by more than V_2.

Next, split the edges into two blocks. Let the first block correspond to edges that change the second coordinate, and the second block correspond to edges that change the first coordinate. Within the first block, order the edges according to their order in $\mathcal{G}_2$ and according to the order of the nodes in the first coordinate. That is, list the edges:

$$(v_1, e_1), ...(v_1, e_{E_2}), (v_2, e_1), ..., (v_2, e_{E_2}), ..., (v_{V_1}, e_1), ..., (v_{V_1}, e_{E_2}).$$

Then the gradient operator can be written as a block matrix:

$$G = \begin{bmatrix} A \\ B \end{bmatrix}. \tag{3.21}$$

where the first block is:

$$A = \begin{bmatrix} G_2 & 0 & ... & 0 \\ 0 & G_2 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & G_2 \end{bmatrix} = I_{V_1} \otimes G_2. \tag{3.22}$$

The second block would take the same form had we ordered the nodes:

$$(1,1), (2,1), ..., (V_1, 1), (1,2), (2,2), ...(V_1, 2), ...(1, V_2), (2, V_2), ...(V_1, V_2).$$

This is precisely the permutation accomplished by the shuffle matrix S_{v_1,v_2}, so:

$$B = (I_{V_2} \otimes G_1) S_{v_1,v_2}. \tag{3.23}$$

Note that the shuffle is applied on the left since rearranging the node indexing changes the columns of the gradient, not the rows.

Therefore the gradient operator is:

$$G = \begin{bmatrix} I_{V_1} \otimes G_2 \\ (I_{V_2} \otimes G_1) S_{v_1,v_2} \end{bmatrix}. \tag{3.24}$$

The node Laplacian follows by evaluating the product $G^\intercal G$ (intermediate steps depend on Equation (3.18))::

$$\begin{aligned} G^\intercal G &= (I_{V_1} \otimes G_2)^\intercal (I_{V_1} \otimes G_2) + S^\intercal_{V_1,V_2} (I_{V_2} \otimes G_1)^\intercal (I_{V_2} \otimes G_1) S_{V_1,V_2} \\ &= (I^\intercal_{V_1} \otimes G_2^\intercal)(I_{V_1} \otimes G_2) + S^\intercal_{V_1,V_2} (I^\intercal_{V_2} \otimes G_1^\intercal)(I_{V_2} \otimes G_1) S_{V_1,V_2} \\ &= (I^\intercal_{V_1} I_{V_1}) \otimes (G_2^\intercal G_2) + S^\intercal_{V_1,V_2} (I^\intercal_{V_2} I_{V_2}) \otimes (G_1^\intercal G_1) S_{V_1,V_2} \\ &= I_{V_1} \otimes (G_2^\intercal G_2) + (G_1^\intercal G_1) \otimes I_{V_2} \end{aligned}.$$

Denote the node Laplacians of the two factor graphs $L^2_{V_1}$ and $L^2_{V_2}$. Then the node Laplacian for the full graph is the Kronecker sum of the node Laplacian on each product graph [40]:

$$L^2_V = L^2_{V_1} \oplus L^2_{V_2} = L^2_{V_1} \otimes I_{V_2} + I_{V_1} \otimes L^2_{V_2} \tag{3.25}$$

This relation allows the node Laplacian for the full graph to be constructed directly from the node Laplacian of the factor graphs. These formulas also generalize easily for the Cartesian products of multiple graphs.[2]

Suppose:

$$\mathcal{G} = \square_{j=1}^n \mathcal{G}_j. \tag{3.28}$$

[2]The gradient of a repeated product can also be computed explicity without iterating Equation (3.24). Reorder the nodes of the product graph so that we count over the first coordinate first, then the second, then

Then the node Laplacian is given by:

$$L_V^2 = \sum_{j=1}^{n} (\otimes_{h=1}^{j-1} I_{V_h}) \otimes L_{V_j}^2 \otimes (\otimes_{h=j+1}^{n} I_{V_h}). \tag{3.29}$$

For example, for the product of three graphs:

$$L_V^2 = L_{V_1}^2 \otimes I_{V_2} \otimes I_{V_3} + I_{V_1} \otimes L_{V_2}^2 \otimes I_{V_3} + I_{V_1} \otimes I_{V_2} \otimes L_{V_3}^2.$$

The Curl and Face Laplacian

The curl is decidedly trickier to construct because the set of loops in $\mathcal{C}_{\text{tree}}$ are not formed from copies of loops that existed in either of the factor graphs. Even so, the structure of the Hammack basis allows the curl to be constructed in blocks using Kronecker products.

As usual, partition the edges according to their factor graphs. Then the curl can be written as the block matrix:

$$C = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \\ & B \end{bmatrix}. \tag{3.30}$$

---

the third, and so on. Then the gradient can be written:

$$G = \begin{bmatrix} (\otimes_{j=2}^{n} I_{V_j}) \otimes G_1 \\ \vdots \\ (\otimes_{h=j+1}^{n} I_{V_h}) \otimes \left((\otimes_{h+1}^{j-1} I_{V_h} \otimes G_j) S_{\prod_{h=1}^{j-1} V_h, V_j} \right) \\ \vdots \\ \left((\otimes_{h=1}^{n-1} I_{V_h}) \otimes G_n \right) S_{\prod_{h=1}^{n-1} V_h, V_n} \end{bmatrix}. \tag{3.26}$$

For example, for a product of four graphs:

$$G = \begin{bmatrix} I_{V_4} \otimes I_{V_3} \otimes I_{V_2} \otimes G_1 \\ I_{V_4} \otimes I_{V_3} \otimes (I_{V_1} \otimes G_2) S_{V_1, V_2} \\ I_{V_4} \otimes (I_{V_1} \otimes I_{V_2} \otimes G_3) S_{V_1 V_2, V_3} \\ (I_{V_1} \otimes I_{V_2} \otimes I_{V_3} \otimes G_4) S_{V_1 V_2 V_3, V_4} \end{bmatrix}. \tag{3.27}$$

The upper block is divided into two diagonal blocks corresponding to the loops in $\mathcal{C}_{\text{factor}}$ that exclusively use edges from $\mathcal{G}_1$, and that exclusively use edges from $\mathcal{G}_2$.

If the edges are ordered lexicographically then:

$$A_1 = I_{V_1} \otimes C_2, A_2 = I_{V_2} \otimes C_1. \tag{3.31}$$

No shuffle matrix is needed since the columns of the curl correspond to the edges, and each block of edges can be ordered according to the scheme introduced for the first block of edges when constructing the gradient.

All that is left is to build B. The matrix B is the curl associated with the Cartesian product of $\mathcal{T}_1$ with $\mathcal{T}_2$ where $\mathcal{T}_1$ and $\mathcal{T}_2$ are spanning trees of the factor graphs. To start we need to introduce an ordering for the loops in this product. Recall that all the loops in the square space are specified by a pair of edges, one from each spanning tree. Thus the loops in the cycle basis for the square space can be indexed by indexing the edges in the spanning trees.

Order the edges in both trees. Then list the pairs of edges:

$$(1,1),(1,2),...(1,V_2-1),(2,1),(2,2),...,(2,V_2-1),...,(V_1,1),(V_1,2),...(V_1,V_2-1).$$

This establishes a definite order for the list of loops in $\mathcal{C}_{\text{tree}}$. To retain consistency with the gradient keep the same ordering of the edges.

Each row of B contains four nonzero elements, two of which correspond to edges in the first factor graph, and two of which correspond to edges in the second factor graph. One approach for constructing B is to build an algorithm which produces the indices of these edges automatically.

To specify the curl we must be able to identify the edge indices of the square in the

product graph corresponding to each pair of edges drawn from the factor graphs. To make sure the algorithm is fast we take advantage of the ordering of the edges to establish a mapping from the value of the fixed coordinate, and the edge in the changing coordinate, to the edge index in $\mathcal{G}$. This first requires a mapping from edges to endpoints.

Denote the two endpoint maps M_1 and M_2. Let $M_1(k) = (i(k), j(k))$ where i, j are the endpoints of edge k in $\mathcal{G}_1$. Define M_2 similarly.

Then, given a loop index in the square space find the corresponding edge pair k, h. From this pair compute the four endpoints $M_1(k), M_2(h) = (i_1, j_1), (i_2, j_2)$. This maps to four edges in $\mathcal{G}$, namely the edges connecting $(i_1, i_2) \to (j_1, i_2) \to (j_1, j_2) \to (i_1, j_2) \to (i_1, i_2)$. The corresponding edge indices, $e_{1,2,3,4}$, in the product graph are:

$$(e_1, e_2, e_3, e_4) = ((i_1-1)E_2+h, (j_1-1)E_2+h, E_2V_1+(i_2-1)E_1+k, E_2V_1+(j_2-1)E_1+k.)$$

Since each edge in the factor graphs must be traversed both forward and backward to traverse a square in the product, the edges can be oriented in $\mathcal{T}_1$ and $\mathcal{T}_2$ so that the corresponding row in B has nonzero elements $1, -1, 1, -1$. This grants a method for building B one row at a time. In psuedocode:

Algorithm for Construction of C ($\mathcal{C}_{\text{tree}}$):

1. pick a spanning tree for each factor graph, $\mathcal{T}_1, \mathcal{T}_2$, and order all edges in factor graphs so that the chords are listed last.
2. Form all pairs (k, h), $k \leq V_1 - 1, h \leq V_2 - 1$ in lexicographic order.
3. Loop over the pairs, and compute endpoints $(i_1, j_1) = M_1(k), (i_2, j_2) = M_2(h)$.
4. Compute edge indices: $(e_1, e_2, e_3, e_4) = ((i_1 - 1)E_2 + h, (j_1 - 1)E_2 + h, E_2V_1 +$

$(i_2 - 1)E_1 + k, E_2V_1 + (j_2 - 1)E_1 + k)$

5. Set the corresponding row in B to sparse$((e_1, e_2, e_3, e_4), (1, -1, 1, -1))$.

Once B is constructed the curl is complete. The face Laplacian can then be computed by taking $CC^\intercal$. The Laplacian can also be written in a block form, and some of the blocks are given by simple Kronecker products. These are the blocks corresponding to the product of A_1 and A_2 with themselves, and take the form $I_{V_1} \otimes L^2_{C_2}$ and $I_{V_2} \otimes L^2_{C_1}$. The rest of the blocks consist of the product of B with $B^\intercal$, and the cross terms between A_1, A_2 and B.

This method can be applied to the gradient and curl of any pair of factor graphs to product the gradient, curl, node Laplacian, and face Laplacian of their Cartesian product. It requires the gradient and curl of the original factor graphs, as well as the edge to endpoint mapping, and a specific choice of spanning trees. The endpoint mapping is usually available directly from the adjacency matrix, which will usually be constructed explicitly in order to build a sparse representation of the gradient. The spanning trees can be built using search procedures, or, if the factor graphs are small, by hand.

An alternative approach is to further subdivide the matrix B into two blocks:

$$B = \begin{bmatrix} B_1 & B_2 \end{bmatrix} \tag{3.32}$$

and express each block as a Kronecker product.

The key idea here is to pay attention to the ordering of the edges and the ordering of the loops. First, the edges are ordered in two blocks: all the edges from the second factor graph, then all edges from the first factor graph. There are V_1E_2 edges in the first set amd V_2E_1 in the second. This division corresponds to the blocking of B. Focus on the first block. Within this block there are V_1 sets of edges with E_2 entries, each corresponding to

the list of all edges in $\mathcal{G}_2$ while the state from $\mathcal{G}_1$ is held fixed. Therefore the columns of B_1 can be can further subdivided into V_1 sets of E_2 columns. Each of these blocks corresponds to a particular node in $\mathcal{G}_1$.

Now consider the ordering of the loops in $\mathcal{C}_{\text{tree}}$. The list of loops is ordered by the lexicographic list of all pairs of edges drawn from the two spanning trees. Therefore the list of loops in $\mathcal{C}_{\text{tree}}$ can be broken into $V_1 - 1$ blocks, each consisting of $V_2 - 1$ rows. Each block corresponds to a particular edge from the first spanning tree, and each row within each block corresponds to a particular edge from the second spanning tree. Either assume that the edges in $\mathcal{G}_1$ and $\mathcal{G}_2$ have been ordered so that the spanning trees correspond to the first $V_1 - 1$ and $V_2 - 1$ entries of each, or let $P(\mathcal{T}_1)$ and $P(\mathcal{T}_2)$ denote permutation matrices, where $P(T_j)$ reorders the edges of $\mathcal{G}_j$ so that the edges in the tree T_j are listed first.

Focus on the first $V_2 - 1$ rows of B_1. These correspond to all the edge pairs from $\mathcal{T}_1$ and $\mathcal{T}_2$ where the edge from $\mathcal{G}_1$ is set to the first edge in $\mathcal{T}_1$. The endpoints of this edge in $\mathcal{G}_1$ are given by the nonzero entries of the corresponding row in the gradient, G_1. These endpoints are the two different values of the first component that appear in the loop. Therefore these endpoints map to blocks in the columns of B_1. These column blocks correspond to the set of all edges from $\mathcal{G}_2$ with the state in $\mathcal{G}_1$ fixed. Since the loops in $\mathcal{C}_{\text{tree}}$ are listed lexicographically in terms of pairs of edges all the first $V_2 - 1$ edges share these two endpoints. Therefore the first $V_2 - 1$ rows of B_1 are zero except in the two blocks of columns corresponding to the endpoints of the first edge in $\mathcal{T}_1$. Moreover, all of the edges in the first column block are crossed in the forward direction, while all of the second column block are crossed in the backwards direction. Therefore the first row block of B_1 is the Kronecker product of the first row of G_1 with some $V_2 - 1 \times E_2$ matrix. Since this logic applies to all the row blocks of B_1 we can infer that B_1 is given by the Kronecker product of the first $V_1 - 1$ rows of G_1 with some $V_2 - 1 \times E_2$ matrix. The first $V_1 - 1$ rows of

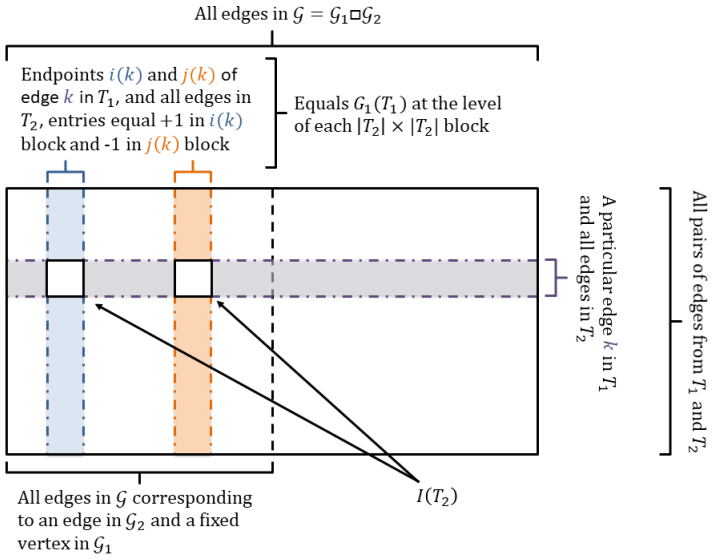

Figure 3.13: Block construction of the component of the curl associated with the square space via a Kronecker product. The large rectangle represents the bottom block of the curl associated with the cycle basis of the square space. It has a row for every pair of edges, one from the spanning tree of each graph. It has a column for every edge in the product graph. The edges in the product graph are broken into two blocks. The first block corresponds to all combinations of a fixed vertex from the first factor graph, and an edge from the second factor graph. These are ordered so that each block of edges corresponds to a particular vertex from the original graph. Therefore, for every edge from the first factor graph there is a block of rows, and two blocks of columns corresponding to the two endpoints of that edge. These blocks are shown in purple, blue, and orange respectively.

G_1 are the restriction of G_1 to the spanning tree $\mathcal{T}_1$. This restriction is accomplished by $I_{V_1-1,E_1} P(\mathcal{T}_1) G_1$ where I_{V_1-1,E_1} is the matrix corresponding to the first $V_1 - 1$ entries of the E_1 by E_1 identity.

Now consider the intersection of the first nonzero row and column block of B_1. The rows corresponds to all loops formed by fixing the edge in $\mathcal{T}_1$ and any edge from $\mathcal{T}_2$, and are listed in order of the edges in $\mathcal{T}_2$. The columns correspond to all edges in $\mathcal{G}$ corresponding to setting the first component to the first endpoint of the edge in $\mathcal{T}_1$ and listing all the edges in $\mathcal{G}_2$. Therefore the k, l entry in this block corresponds to an edge in a loop in the square space of $\mathcal{G}$ formed by the chosen edge in $\mathcal{T}_1$ and the k^{th} edge in $\mathcal{T}_2$. The corresponding

edge in $\mathcal{G}$ is given by setting the first component to the first endpoint of the chosen edge in $\mathcal{T}_1$ and crossing the l^{th} edge in $\mathcal{G}_2$. The loop only includes the edge if the l^{th} edge in $\mathcal{G}_2$ corresponds to the k^{th} edge in $\mathcal{T}_2$. If the edges are listed so that all the edges in the tree appear first this means that this entry is nonzero if and only if $k = l$. Therefore, the first nonzero block of B_1 is I_{V_2-1,E_2} if the edges are listed so that the all edges in the tree appear first. Otherwise we need to permute the edge ordering, so the first nonzero block is $I(V_2 - 1, E_2)P(T_2)^\intercal$. This same logic applies to all the nonzero blocks of B_1 so:

$$B_1 = (I_{V_1-1,E_1}P(T_1)G_1) \otimes (I(V_2 - 1, E_2)P(T_2)^\intercal). \tag{3.33}$$

The block structure of B_1 is illustrated in Figure 3.13.

The same construction would apply to B_2 if the loops were listed by walking through the edges in $\mathcal{T}_1$ before the edges in $\mathcal{T}_2$. Therefore the second block is given by the same construction as the first block, only with the rows shuffled to match the alternate ordering of the loops:

$$B_2 = S_{V_1-1,V_2-1}\left[(I_{V_2-2,E_1}P(\mathcal{T}_2)G_2) \otimes (I(V_1 - 1, E_1)P(\mathcal{T}_1)^\intercal)\right]. \tag{3.34}$$

Let $G_1(\mathcal{T}_1) = I_{V_1-1,E_1}P(\mathcal{T}_1)G_1$ and $G_2(\mathcal{T}_2) = I_{V_2-2,E_1}P(\mathcal{T}_2)G_2$ denote the gradients of each spanning tree. Similarly let $I(\mathcal{T}_1) = I(V_1 - 1, E_1)P(\mathcal{T}_1)^\intercal$ and $I(\mathcal{T}_2) = I(V_2 - 1, E_2)P(\mathcal{T}_2)^\intercal$. Then the curl of $\mathcal{C}_{\text{tree}}$ is:

$$C = \begin{bmatrix} I_{V_1} \otimes C_2 & 0 \\ 0 & I_{V_2} \otimes C_1 \\ G_1(\mathcal{T}_1) \otimes I(\mathcal{T}_2) & S_{V_1-1,V_2-1}\left[G_2(\mathcal{T}_2) \otimes I(\mathcal{T}_1)\right] \end{bmatrix}. \tag{3.35}$$

Equation (3.35) gives a more direct method for computing the curl than the algorithmic method introduced before.

To find the face Laplacian take the product of the curl with the adjoint curl. This is:

$$L^2_{\mathcal{C}} = \begin{bmatrix} I_{V_1} \otimes L^2_{\mathcal{C}_2} & 0 & (G_1(\mathcal{T}_1) \otimes C(\mathcal{T}_2)^\intercal)^\intercal \\ 0 & I_{V_2} \otimes L^2_{\mathcal{C}_1} & (G_2(\mathcal{T}_2) \otimes C_1(\mathcal{T}_1))^\intercal S^\intercal_{V_1-1,V_2-1} \\ G_1(\mathcal{T}_1) \otimes C(\mathcal{T}_2)^\intercal & S_{V_1-1,V_2-1}(G_2(\mathcal{T}_2) \otimes C_1(\mathcal{T}_1)) & G_1(\mathcal{T}_1)G_1^\intercal(\mathcal{T}_1) \oplus G_2(\mathcal{T}_2)G_2^\intercal(\mathcal{T}_\in) \end{bmatrix} \tag{3.36}$$

where $(\mathcal{T}_j)$ denotes the restriction of the operator to the edges in $\mathcal{T}_j$. For the node Laplacians this means the node Laplacian of the spanning trees. Notice that the first two diagonal blocks are built from the face Laplacians of the factor graphs, while the remaining blocks are products of the restricted gradient and curl. Also notice that the bottom block has the same form as the node Laplacian for the Cartesian product of the spanning trees, except that it is the Kronecker of the gradient of the divegrence, not of the divergence of the gradient. The divergence of the gradient is the node Laplacian. The gradient of the divergence is related to the graph Helmholtzian [15, 16], and is closely related to the signed adjacency matrix of the edge/line graph (see Section 4.6). This result generalizes the observation that $\nabla \times \nabla = \nabla(\nabla\cdot) - \nabla \cdot \nabla$ in $\mathbb{R}^3$ to the Cartesian product of trees.

The next section will take advantage of the structure of the operators to introduce intuitive spectral methods for solving the discrete Poisson equation. This analysis will include an explicit construction of the eigenvalues and eigenvectors in terms of the eigenvalues and eigenvectors of the component Laplacians. Expressing the spectrum of the product Laplacain in terms of the spectrum of the product Laplacians will lead to a solution method analogous to separation of variables, and will enable an explicit algorithm for solving the discrete Poisson equation on lattices that uses the Fast Fourier Transform (FFT) [70] to transform into and out of the eigenbasis of the node Laplacian efficiently.

3.3.3 Cartesian Products: Laplacian and Spectral Methods

In the previous section we introduced methods for building the operators associated with the HHD of the Cartesian product of two graphs. This section will focus on an analytical method for solving the discrete Poisson equation associated with scalar potential. Once the scalar potential is found the conservative flow can be recovered from the gradient of the scalar potential and the rotational flow can be recovered by subtracting the conservative flow from the full edge flow. If a fundamental cycle basis is used on each factor graph then the Hammack basis is a fundamental cycle basis for the product graph, so the rotational potential can be recovered directly from the rotational flow without any need to solve a linear system.

The tool introduced here for solving the discrete Poisson equation is a separation of variables approach inspired by the spectral solution to the Lyapunov equation (see [71]). The eigenvalues and eigenvectors of the node Laplacain $\mathcal{L}$ can be built explicitly from the eigenvalue decomposition of the factor graphs [61]. This provides an intuitive spectral method for solving the Poisson equation.

Consider the discrete Poisson equation associated with the product of two graphs:

$$L^2_{\mathcal{V}}\phi = -G^{\intercal} f, \quad L^2_{\mathcal{V}} = (G_1^{\intercal} G_1) \oplus (G_2^{\intercal} G_2). \tag{3.37}$$

The potential ϕ is defined on all $V_1 V_2$ nodes in the product. Define the matrix Φ whose i, j entry is $\phi_{i,j}$ where (i, j) indexes a pair of nodes from $\mathcal{G}_1$ and $\mathcal{G}_2$. Then $\phi = \text{vec}(\Phi)$. The discrete Poisson equation can then be recast as a matrix equation using the identity [69]:

$$(A \otimes B)\text{vec}(M) = \text{vec}(AMB). \tag{3.38}$$

Then:

$$G^\intercal G\phi = \left[(G_1^\intercal G_1) \otimes I_{V_2}\right]\phi + \left[I_{V_1} \otimes (G_2^\intercal G_2)\right]\phi = \text{vec}\left(G_1^\intercal G_1 \Phi I_{V_2} + I_{V_1}\Phi G_2^\intercal G_2\right).$$

Multiplication by the identities does not change Φ, so, if mat() denotes the inverse operation to vec(), then the discrete Poisson equation can be written as the matrix equation:

$$G_1^\intercal G_1 \Phi + \Phi G_2^\intercal G_2 = -\text{mat}(G^\intercal f). \tag{3.39}$$

Let $D_f = -\text{mat}(G^\intercal f)$ be the matrix equal to the divergence of the edge flow. Then the matrix Poisson equation reduces to:

$$L_{\mathcal{V}_1}^2 \Phi + \Phi {L_{\mathcal{V}_2}^2}^\intercal = D_f. \tag{3.40}$$

The transpose on the second Laplacian is introduced for convenience. Since both Laplacians are symmetric it makes no difference whether they are transposed. In this form the matrix Poisson equation is an example of the Sylvester equation [69, 72]:

$$AX + XB^\intercal = C \tag{3.41}$$

and is similar in form to the Lyapunov equation:

$$AX + XA^\intercal = BB^\intercal. \tag{3.42}$$

The Sylvester equation is important in control theory [69], and is used to approximate the solution to the Poisson equation on rectangular domains using finite differences [73]. The Lyapunov equation is a special case of the Sylvester equation, and plays a central role the

analysis of stochastic processes.

The Sylvester equation admits an elegant solution analogous to separation of variables. It can be solved efficiently by a Schur factorization [74], or by spectral methods. Efficient solution methods are addressed at length in [72]. Our solution leverages the relation between the spectrum of the two component Laplacians and the Laplacian of their Kronecker sum. Spectral properties of Kronecker products and sums are used widely in modal analysis of stiffness matrices, and other matrices related to regular graph structures [61].

Expand both component Laplacians. Since both are real symmetric both are unitarily diagonalizable:

$$\begin{aligned} L_{\mathcal{V}_1}^2 &= U\Lambda U^\intercal \\ L_{\mathcal{V}_2}^2 &= W\Sigma W^\intercal \end{aligned} \tag{3.43}$$

where U and W are orthonormal matrices.

Then the solution can be expressed in three steps:

$$\hat{D}_f = U^\intercal \text{mat}(G^\intercal f) W^\intercal \quad \rightarrow \quad \hat{\phi}_{ij} = -\frac{\hat{d}_{ij}}{\lambda_i + \sigma_j}, \hat{\phi}_{11} = 0 \quad \rightarrow \quad \Phi = U\hat{\Phi}W. \tag{3.44}$$

Notice the similarity to the spectral solution to a system of linear equations. The first step amounts to transforming the right hand side into the eigenbasis. The second step amounts to dividing by the nonzero eigenvalues. And the third step amounts to transforming back onto the original basis.[3]

[3]The solution to the Sylvester equation given in Equation (3.44) can be guessed using an ansatz inspired by the solution of the Lyapunov equation. The Lyapunov equation is the steady state equation for the covariance of stochastic processes with linear rates. It is solved [71, 75] by:

$$X = \int_0^\infty \exp(As) B B^\intercal \exp(A^\intercal s) ds \tag{3.45}$$

where the integral over s comes from the long time limit of a process whose dynamics are driven by the matrix A and perturbed by a noise source characterized by B.

Therefore we take the following integral as an ansatz (where the negatives are introduced by multiplying

In fact, Equation (3.44) is the spectral solution to the discrete Poisson equation in terms of the spectrum of the node Laplacian of the product graphs. To show that the proposed solution is the spectral solution we find the eigen-decomposition of $L^2_{\mathcal{V}}$ in terms of the eigenvalues and eigenvectors of the factor graphs, then solve the discrete Poisson equation using the spectral method.

The key first step is analogous to separation of variables in the continuum.

---

the discrete Poisson equation by -1 on both sides and ensure that the integrals converge):

$$\Phi = -\int_0^\infty \exp\left(-L^2_{\mathcal{V}_1} s\right) D_f \exp\left(-{L^2_{\mathcal{V}_2}}^\intercal s\right) ds. \tag{3.46}$$

We can use the eigenvalue decompositions of the node Laplacians to check that Φ satisfies the matrix Poisson equation. First, expand both Laplacians in the integral. Then the integral is:

$$-\int_0^\infty \exp\left(-L^2_{\mathcal{V}_1} s\right) D_f \exp\left(-{L^2_{\mathcal{V}_2}}^\intercal s\right) ds = U \int_0^\infty \exp\left(-\Lambda s\right) U^\intercal D_f W^\intercal \exp\left(-\Sigma s\right) ds W.$$

Let $\hat{D}_f = U^\intercal D_f W^\intercal$. Then, since Λ and Σ are diagonal, the i, j entry of the integral is:

$$-\int_0^\infty \hat{d}_{ij} \exp\left(-(\lambda_i + \sigma_j) s\right) ds.$$

All of the eigenvalues of $L^2_{\mathcal{V}_1}$ and $L^2_{\mathcal{V}_2}$ are positive and real except for the first eigenvalue of each, which is zero. The zero eigenvalues correspond to eigenvectors with constant entries (proportional to $\mathbf{1}$). So, if Φ is chosen so that the mean value of ϕ is zero, then the any projection onto $\mathbf{1}$ is zero by convention. Then the remaining integrals all converge to:

$$-\int_0^\infty \hat{d}_{ij} \exp\left(-(\lambda_i + \sigma_j) s\right) ds = \frac{\hat{d}_{ij}}{\lambda_i + \sigma_j} \quad i \neq 1 \text{ or } j \neq 1.$$

So:

$$\hat{\Phi}_{ij} = \left\{ \begin{array}{cc} 0 & \text{if } i = 1 \text{ and } j = 1 \\ \hat{d}_{ij} / (\lambda_i + \sigma_j) & \text{else} \end{array} \right\}, \Phi = U \hat{\Phi} W.$$

Substituting into the discrete Poisson equation:

$$L^2_{\mathcal{V}_1} \Phi + \Phi {L^2_{\mathcal{V}_2}}^\intercal = U \Lambda U^\intercal U \hat{\Phi} W + U \hat{\Phi} W W^\intercal \Sigma W = -U \left[\Lambda \hat{\Phi} + \hat{\Phi} \Sigma\right] W$$

But:

$$\left[\Lambda \hat{\Phi} + \hat{\Phi} \Sigma\right]_{ij} = \lambda_i \frac{\hat{d}_{ij}}{\lambda_i + \sigma_j} + \sigma_j \frac{\hat{d}_{ij}}{\lambda_i + \sigma_j} = \hat{d}_{ij}$$

so:

$$L^2_{\mathcal{V}_1} \Phi + \Phi {L^2_{\mathcal{V}_2}}^\intercal = U \hat{D}_f W = D_f = -\text{mat}(G^\intercal f).$$

Therefore the integral solution to the Lyapunov equation also provides an explicit solution to the discrete Poisson equation in terms of the eigenvalue decomposition of the component Laplacians.

Let U be the eigenvectors of $L^2_{\mathcal{V}_1}$ and W be the eigenvectors of $L^2_{\mathcal{V}_2}$ with eigenvalues Λ and Σ respectively. Then consider the matrix $V^{h,k}$ defined by the outer product $u_h w_k^\intercal$. This outer product is analogous to separation of variables since the i,j entry of the matrix $v^{j,k}_{i,j} = u_{ih} w_{jk}$. Then the product:

$$\begin{aligned} L^2_{\mathcal{V}} \text{vec}(u_h w_k^\intercal) &= \text{vec}(L^2_{\mathcal{V}_1} u_h w_k^\intercal + u_h w_k^\intercal {L^2_{\mathcal{V}_2}}^\intercal) = \text{vec}(\lambda_h u_h w_k^\intercal + u_h w_k^\intercal \sigma_k) \\ &= (\lambda_h + \sigma_k) \text{vec}(u_h w_k^\intercal). \end{aligned} \tag{3.47}$$

Therefore the outer product of any two eigenvectors of $L^2_{\mathcal{V}_1}$ and $L^2_{\mathcal{V}_2}$ corresponds to an eigenvector of $L^2_{\mathcal{V}}$ with eigenvalue given by the sum of the eigenvalues of the two eigenvectors of the component Laplacians. Since both component Laplacians are real symmetric they are unitarily diagonalizable, so each have a set of V_1, or V_2, distinct eigenvectors. Then there are $V_1 V_2$ distinct matrices which can be formed by outer-products of the sets of eigenvectors. These outer products form an orthonormal basis for the space of $V_1 V_2$ by $V_1 V_2$ matrices, since the matrix inner product $\langle u_h w_k^\intercal, u_l w_n^\intercal \rangle = \sum_{i,j} u_{ih} w_{kj} u_{il} w_{nl} = (u_h^\intercal u_l)(w_k^\intercal w_n) = \delta_{hl} \delta_{kn}$. The Laplacian, $L^2_{\mathcal{V}}$ is square with dimension $V_1 V_2$. Therefore all eigenvectors of $L^2_{\mathcal{V}}$ correspond to outer-products of eigenvectors of the factor graphs. It follows that the entire spectrum of $L^2_{\mathcal{V}}$ can be built directly from the spectra of the factor graphs. Since $L^2_{\mathcal{V}}$ is itself real symmetric this eigenbasis is orthonormal.

Let $v^{h,k} = \text{vec}(u_h w_k^\intercal)$ denote the h,k eigenvector of $L^2_{\mathcal{V}}$, with corresponding eigenvalue $\mu^{h,k} = \lambda_h + \sigma_k$. The eigenvectors can be stored in a matrix by evaluating the Kronecker product of the matrices of eigenvectors of the component Laplacians, $W \otimes U$. The discrete Poisson equation can then be solved by expanding the right hand side in the eigenbasis formed by outer products of the eigenvectors of the components, rescaling by the eigenvalues, then transforming back to the original basis. Since the eigenbasis is orthonormal the

expansion of $D_f = -\text{mat}(G^\intercal f)$ onto the eigenbasis only requires the inner product:

$$\hat{d}_{h,k} = \sum_{i=1}^{V_1}\sum_{j=1}^{V_2} u_{ih} w_{jk} \text{mat}(-G^\intercal f)_{i,j} = -u_h^\intercal \text{mat}(G^\intercal f) w_k. \tag{3.48}$$

Notice that, since $u_1 \propto \mathbf{1}$ and $v_1 \propto \mathbf{1}$, and $\mathbf{1}^\intercal G^\intercal = (G\mathbf{1})^\intercal = \mathbf{0}^\intercal$, the one-one entry $\hat{d}^{1,1} = 0$.

The corresponding potential is recovered by scaling the coefficients in the expansion by the sums of the eigenvalues, and mapping out of the eigenbasis:

$$\Phi = \sum_{h=1}^{V_1}\sum_{k=1}^{V_2} \frac{\hat{d}^{h,k}}{\lambda_h + \sigma_k} u_h w_k^\intercal. \tag{3.49}$$

The structure of the spectrum of the product Laplacian makes it possible to perform the decomposition onto the eigenvectors one factor graph at a time. In general, the coefficients of the decomposition are a double sum over the indexes i, j (see Equation (3.48)). Therefore, if we define the coefficients $\hat{b}^{h,j}$ and $\hat{c}^{i,k}$:

$$\hat{b}^{h,j} = \sum_{i=1}^{V_1} u_h(i)[G^\intercal f]_{i,j}$$
$$\hat{c}^{i,k} = \sum_{j=1}^{V_2} w_k(j)[G^\intercal f]_{i,j}.$$

Then, by rearranging the order of the sum:

$$\hat{d}^{h,k} = \sum_{j=1}^{V_1} w_k(j)\hat{b}^{h,j} = \sum_{i=1}^{V_2} u_h(i)\hat{c}^{i,k}. \tag{3.50}$$

The first approach considers each column of $\text{mat}(G^\intercal f)$ independently, finds the coefficients of the eigenvector expansion of each column using the eigenvalues of $L^2_{\mathcal{V}_1}$, then expands each row of coefficients $\hat{b}^{h,j}$ into the eigenvector basis of $L^2_{\mathcal{V}_2}$. The second ap-

proach is the same, but with the expansion performed on the rows before the columns. Separating the operations one component at a time is attractive since in some important cases the expansion onto the eigenbasis associated with each product graph may be done efficiently. In particular, when the factor graphs are rings or lines then the decomposition step can be performed using FFT. In those cases the spectral approach is equivalent to performing a sequence of FFT's [72].

Now suppose that $\mathcal{G} = \square_{j=1}^{d} \mathcal{G}_j$. Then the product Laplacian, $L^2_{\mathcal{V}}$ has $\prod_{j=1}^{d} V_j$ eigenvectors of the form:

$$v(x) = \prod_{j=1}^{d} u_j(x_j) \tag{3.51}$$

where x is an index vector, $x_j \in [1, 2, ..., V_j]$ where V_j is the number of vertices in $\mathcal{G}_j$, and u_j is the j^{th} eigenvector of $L^2_{\mathcal{V}_j}$. The corresponding eigenvalues are:

$$\mu(x) = \sum_{j=1}^{d} \lambda_j(x_j) \tag{3.52}$$

where λ_j are the eigenvalues of $L_{V,j}$. This expansion can be checked by iteratively applying the expansion for the product of pairs of graphs.

This eigenvalue expansion is a powerful tool for understanding the spectrum of Hamming graphs, generalized hypercubes, and other large Cartesian products. If the factor graphs are small, or familiar, then the spectrum of the product Laplacian is easily understood as products of the eigenvectors of the component Laplacians, with eigenvalues equal to the sums of the eigenvalues of the components. For the two most important factor graphs, rings and paths, the eigenvectors are trigonometric functions, so the corresponding expansion can be accomplished via an Fast Fourier Transform (FFT). The solution via the FFT mimics the separation of variables solution to the Poisson equation in the continuum

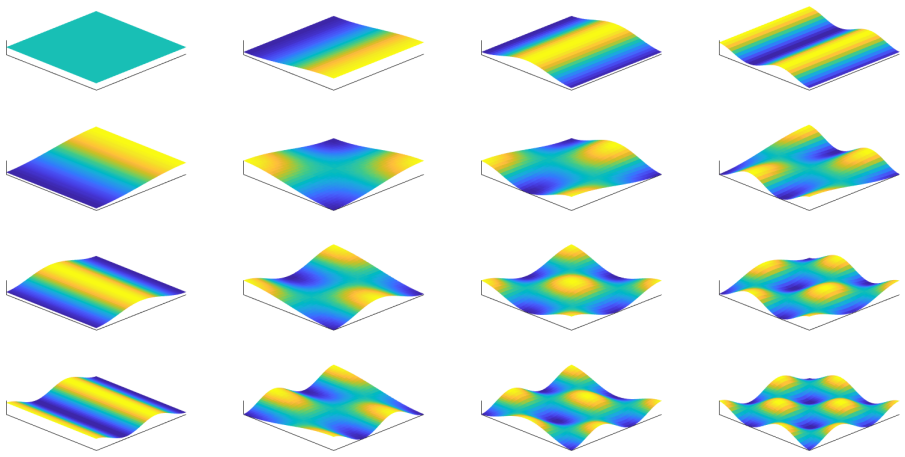

Figure 3.14: The first 16 eigenmodes of a 30 by 60 lattice. Notice that the eigenmodes are all products of the lower eigenmodes along the top row, and first column.

which relies on the Fourier transform. The first 16 eigenmodes of a 30 by 60 grid are shown in Figure 3.14 as an example. Notice that all the eigenmodes i, j are the product of the $i, 1$ and $1, j$ eigenmodes, since each eigenmode is an outer product of the eigenvectors of each Laplacian.

In Section 3.4 this spectral approach is applied to derive FFT based methods for solving the discrete Poisson equation on Cartesian products of rings and lines. These include a variety of important special cases including grids and lattices with or without periodic boundaries, and hypercubes.

3.4 Grids, Lattices, and Hypercubes

In Section 3.3 we showed that, if the graph of interest can be expressed as the Cartesian product of a set of factor graphs, then the discrete Poisson equation defining the scalar potential can be solved using the eigenvectors and eigenvalues of the node Laplacians of

the factor graphs. It follows that solutions to the discrete Poisson equation for a variety of product graphs can be understood directly from the spectra of their factor graphs. In Section 3.4.1 we present the spectrum of three important factor graphs: complete graphs, lines, and cycles. Building from these factor graphs, we consider hypercubes and lattices (with and without periodic boundaries) since, in these cases, the expansion into the eigenbasis can be performed efficiently with either a fast Hadamard transform or an FFT.

3.4.1 Spectra of Important Factor Graphs

Complete Graphs

The node Laplacian for a complete graph with V nodes equals $VI - \mathbf{1}\mathbf{1}^\intercal$ where $\mathbf{1}$ is the vector of all ones (see Section 3.2.4). Therefore the first eigenvector of $L^2_{\mathcal{V}}$ is $\mathbf{1}$ with eigenvalue 0. Any vector orthogonal to $\mathbf{1}$ is also an eigenvector of $L^2_{\mathcal{V}}$ with eigenvalue equal to the number of vertices in the graph, V. Let $Q \in \mathbb{R}^{V \times V-1}$ be a matrix with orthonormal columns, all orthogonal to $\mathbf{1}$. Then the unitary matrix $\left[\sqrt{\frac{1}{V}}\mathbf{1}|Q\right]$ diagonalizes $L^2_{\mathcal{V}}$, and the corresponding eigenvalues are $[0, V, V, ...V]$.

Lines

Consider a graph consisting of V nodes connected in a line. If $V = 2$ then the graph is complete, so has eigenvalues 0 and $V = 2$ corresponding to eigenvectors $\frac{1}{\sqrt{2}}\mathbf{1} = \frac{1}{\sqrt{2}}[1; 1]$ and $\frac{1}{\sqrt{2}}[1; -1]$.

If $V > 2$ then $L^2_{\mathcal{V}}$ equals:

$$L^2_{\mathcal{V}} = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix}_{V \times V}$$

Note that this matrix is the same as the standard second-order central difference operator with Neumann boundary conditions and discretization length equal to one, all multiplied by -1 [76]. The spectrum of this matrix is well known. As in the continuum, the eigenvectors are trigonometric functions. The eigenvectors and eigenvalues of the node Laplacian are:

$$\begin{aligned} \lambda_j &= 4\sin^2\left(\frac{\pi(j-1)}{2V}\right) \\ v_{ij} &= \left\{ \begin{array}{l} V^{-1/2} \text{ if } j = 1 \\ \sqrt{\frac{2}{V}}\cos\left(\frac{\pi(j-1)}{V}\left(i - \frac{1}{2}\right)\right) \text{ else} \end{array} \right\}. \end{aligned} \tag{3.53}$$

Cycles

Consider a graph consisting of V nodes forming a cycle. Then $L_{\mathcal{V}}^2$ equals:

$$L_{\mathcal{V}}^2 = \begin{bmatrix} 2 & -1 & & & & -1 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ -1 & & & & -1 & 2 \end{bmatrix}_{V \times V}$$

Note that this matrix is the same as the standard second-order central difference operator with periodic boundary conditions and discretization length equal to one, all multiplied by -1 [76]. The spectrum of this matrix is well known. As before, the eigenvectors are trigonometric functions. The eigenvectors and eigenvalues of the node Laplacian are:

$$\begin{aligned} \lambda_j &= \left\{ \begin{array}{l} 4\sin^2\left(\dfrac{\pi(j-1)}{2V}\right) \text{ if } j \text{ odd} \\ 4\sin^2\left(\dfrac{\pi j}{2V}\right) \text{ if } j \text{ even} \end{array} \right\} \\ v_{ij} &= \left\{ \begin{array}{l} V^{-1/2} \text{ if } j = 1 \\ \sqrt{\dfrac{2}{V}}\sin\left(\dfrac{\pi j}{V}\left(i - \dfrac{1}{2}\right)\right) \text{ if } j \in [2, V-1] \text{ and is even} \\ \sqrt{\dfrac{2}{V}}\cos\left(\dfrac{\pi(j-1)}{V}\left(i - \dfrac{1}{2}\right)\right) \text{ if } j \in [2, V] \text{ and is odd} \\ V^{-1/2}(-1)^i \text{ if } j = V \text{ and } V \text{ is even} \end{array} \right\}. \end{aligned} \tag{3.54}$$

3.4.2 Hypercubes

The simplest set of Cartesian product graphs to consider are hypercubes. In this section we apply the spectral method developed in Section 3.3.3 to the special case of hypercubes. The spectral method can be performed in $V\log(V)$ operations on a hypercube using the fast Walsh-Hadamard transform. The Hadamard transform is well studied, particularly in image processing, and is equivalent to the $2 \times 2 \times ...2$ FFT [77, 78, 79]. This efficient implementation makes the HHD fast for high-dimensional hypercubes when a direct implementation of the spectral method would be too slow.

A d dimensional hypercube is the iterated Cartesian product of the line graph with two nodes. It has 2^d vertices, and its nodes can be represented as d bit strings. Since hypercubes can be formed by repeated Cartesian products, the spectrum of the node Laplacian for any hypercube can be constructed from the spectrum of the node Laplacian for a line graph with two vertices. A line graph with two vertices is complete, so has eigenvalues 0 and 2. The null-vector is parallel to $[1, 1]$, so the remaining eigenvector must be parallel to $[1, -1]$ since eigenvectors of symmetric matrices are orthogonal. Define the symmetric matrix:

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{3.55}$$

The matrix H_1 is a matrix of unnormalized eigenvectors for the node Laplacian of each factor graph. Then, the unnormalized eigenvectors for the 2 dimensional hypercube are the columns of the matrix $H_2 = H_1 \otimes H_1$. The unnormalized eigenvectors of the d dimensional hypercube are given by the recursion (see Section 3.3.3):

$$H_d = H_1 \otimes H_{d-1}, \quad H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{3.56}$$

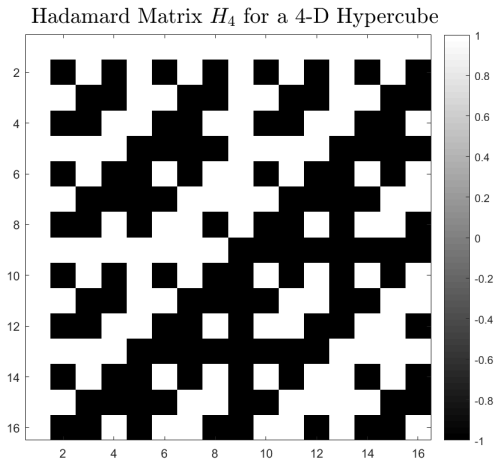

Figure 3.15: Hadamard matrix H_4 for a 4 dimensional hypercube. White entries equal one and black entries equal negative one.

The matrix H_d is the d^{th} Hadamard matrix as originally introduced by Sylvester [77, 78]. Hadamard matrices are all symmetric. [4] The Hadamard matrices are also orthogonal [77] since the columns of H_d are eigenvectors corresponding to the node Laplacian, which is symmetric. Therefore, $\frac{1}{2^{d/2}} H_d = \left(\frac{1}{2^{d/2}} H_d\right)^{-1}$. To recover the normalized eigenvectors simply divide by $2^{d/2}$:

$$L_{\mathcal{V}}^2 = 2^{-d} H_d \Lambda H_d. \tag{3.57}$$

An 4 dimensional example is shown in Figure 3.15. Notice that the rows form square wave patterns. Since H_d is orthogonal, multiplication by H_d is equivalent to expansion on a square wave basis. Expansion onto this basis is a Walsh transform [80].

For low dimensional hypercubes the Hadamard matrices can be visualized as shown in Figure 3.16. Each cube represents a row of the matrix, where white nodes represent entries with value 1, and red nodes represent entries with value -1.

[4] The symmetry of Hadamard matrices can be shown by induction. The first Hadamard matrix, H_1, is symmetric. To extend to arbitrary d use the induction hypothesis $H_{d-1}^\intercal = H_{d-1}$ then $H_d^\intercal = (H_1 \otimes H_{d-1})^\intercal = H_1^\intercal \otimes H_{d-1}^\intercal = H_1 \otimes H_{d-1} = H_d$. proving that H_d is symmetric for any d.

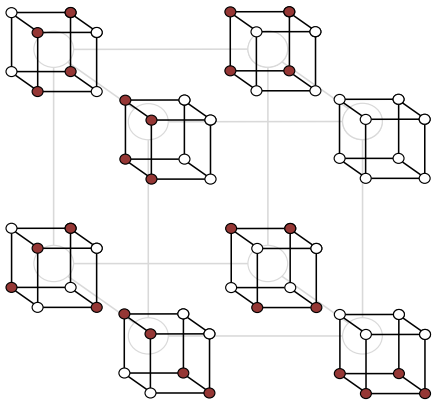

Figure 3.16: The Hadamard matrix for a cube represented by 8 copies of the cube, each corresponding to a row in H_3. The white node represent entries with $+1$ and the red nodes represent entries with -1.

The rows of the Hadamard matrix H_d can be uniquely associated with a d bit string. The strings are ordered lexicographically so that $000...00$ is followed by $000...01$ then $000...10$, then $000...11$ and so on. Therefore the j^{th} row maps to the d bit string representing $j-1$ base 2. The corresponding eigenvalues are given by the number of nonzero entries in the base two representation of $j-1$. Let $\text{base}_2(j-1,d)$ be the base two representation of $j-1$ using d bit strings. Then the corresponding eigenvalue is:

$$\lambda_j = 2|\text{base}_2(j-1,d)| \tag{3.58}$$

where $|\text{base}_2(j-1,d)|$ is the number of nonzero entries in the d bit string.

Then, to implement the spectral method directly:

1. Compute $G^\intercal f$
2. Compute $2^{-d/2}H_d(G^\intercal f)$
3. Compute $\Lambda^\dagger(2^{-d/2}H_d G^\intercal f)$

4. Compute $2^{-d} H_d(\Lambda^\dagger H_d G^\intercal f)$.

If applied directly the products with H_d require $(2^d)^2$ operations each, and the rescaling by the eigenvalues requires 2^d operations. Therefore direct application of the spectral method would require $\mathcal{O}(V^2) = \mathcal{O}(2^{2d})$ operations. This can be dramatically improved by taking advantage of the recursive structure of the Hadamard matrices.

Consider H_2, the Hadamard matrix for a square:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Multiplication of an arbitrary vector v by H_2 requires computing:

$$v_{00} + v_{01} + v_{10} + v_{11}$$

$$v_{00} - v_{01} + v_{10} - v_{11}$$

$$v_{00} + v_{01} - v_{10} - v_{11}$$

$$v_{00} - v_{01} - v_{10} + v_{11}.$$

Performed directly this requires $12 = V(V-1)$ additions.

Instead, compute $w_{0+} = v_{00} + v_{01}, w_{0-} = v_0 - v_{01}, w_{1+} = v_{10} + v_{11}, w_{1-} = v_{10} - v_{11}$.

This requires 4 additions. Then, to compute $H_2 v$ we only need to perform:

$$w_{0+} + w_{1+}$$

$$w_{0-} + w_{1-}$$

$$w_{0+} - w_{1+}$$

$$w_{0-} - w_{1-}$$

which also only requires 4 operations. Computing w first reduces the total cost to $8 = V \log_2(V)$ operations instead of 12. The same approach can be easily scaled up to a cube. Multiplication by H_3 requires:

$$v_{000} + v_{001} + v_{010} + v_{011} + v_{100} + v_{101} + v_{110} + v_{111}$$

$$v_{000} - v_{001} + v_{010} - v_{011} + v_{100} - v_{101} + v_{110} - v_{111}$$

$$v_{000} + v_{001} - v_{010} - v_{011} + v_{100} + v_{101} - v_{110} - v_{111}$$

$$v_{000} - v_{001} - v_{010} + v_{011} - v_{100} - v_{101} + v_{110} - v_{111}$$

$$v_{000} + v_{001} + v_{010} + v_{011} - v_{100} - v_{101} - v_{110} - v_{111}$$

$$v_{000} - v_{001} + v_{010} - v_{011} - v_{100} + v_{101} - v_{110} + v_{111}$$

$$v_{000} + v_{001} - v_{010} - v_{011} - v_{100} - v_{101} + v_{110} + v_{111}$$

$$v_{000} - v_{001} - v_{010} + v_{011} - v_{100} + v_{101} + v_{110} - v_{111}.$$

If done directly this sum requires $8 \times 7 = V(V-1) = 56$ operations. To streamline

the product compute:

$$
\begin{aligned}
w_{00+} &= v_{000} + v_{001}, & w_{00-} &= v_{000} - v_{001} \\
w_{01+} &= v_{010} + v_{011}, & w_{01-} &= v_{010} - v_{011} \\
w_{10+} &= v_{100} + v_{101}, & w_{10-} &= v_{100} - v_{101} \\
w_{11+} &= v_{110} + v_{111}, & w_{11-} &= v_{110} - v_{111}
\end{aligned}
$$

Computing w requires $8 = V$ computations. Now the original sum has the form:

$$
\begin{aligned}
& w_{00+} + w_{01+} + w_{10+} + w_{11+} \\
& w_{00-} + w_{01-} + w_{10-} + w_{11-} \\
& w_{00+} - w_{01+} + w_{10+} - w_{11+} \\
& w_{00-} - w_{01-} + w_{10-} - w_{1-} \\
& w_{00+} + w_{01+} - w_{10+} - w_{11+} \\
& w_{00-} + w_{01-} - w_{10-} - w_{11-} \\
& w_{00+} - w_{01+} - w_{10+} + w_{11+} \\
& w_{00-} - w_{01-} - w_{10-} + w_{11-}.
\end{aligned}
$$

Applied directly this requires 24 computations, however this is just H_2 applied to w_+ and w_-. As shown before, H_2 can be applied with only 8 computations, so, using the efficient implementation of H_2, computing $H_2 w_+$ and $H_2 w_-$ only requires 16 computations. Therefore the total cost of applying H_3 is only $24 = (2^3)3 = V \log_2(V)$ computations, not $V(V-1) = 56$.

This technique can be applied recursively to perform the product with any Hadamard matrix H_d. The recursive algorithm is the fast Walsh-Hadamard Transform [79].The trans-

form only requires $V \log_2(V) = d2^d$ computations since there are d iterations, and each requires 2^d computations. The transform is implemented in Matlab by the fwht command.

Thus, by using the fast Walsh-Hadamard transform (fWHt) to map into and out of the eigenbasis of the Laplacian, the spectral method for solving the discrete Poisson equation can be performed efficiently. Using the fWHt to perform products with Hadamard matrices reduces the computation cost to order $d2^d$ instead of 2^{2d} or, worse, 2^{3d} computations (for direct inversion of the Laplacian).

3.4.3 Lattices

Let $\mathcal{G} = \square_{j=1}^{d} \mathcal{G}_j$ be the product graph formed by the repeated Cartesian product of a sequence of factor graphs, $\mathcal{G}_j$, which are all either lines or cycles. Then $\mathcal{G}$ is a d-dimensional lattice, with periodic boundaries in the dimension corresponding to factor graphs that are cycles. The node Laplacian of the product graph is $L^2_{\mathcal{V}} = \oplus_{j=1}^{d} L^2_{\mathcal{V}_j}$ so the spectrum of the node Laplacian can be constructed from the spectrum of the Laplacian on each factor graph. Since all of the factors are either lines or loops the eigenvectors of the factor graphs are all trigonometric functions (see Section 3.4.1).

Since the spectrum of the node Laplacian is necessarily orthonormal, moving into the eigenbasis of the node Laplacian requires taking an inner product with each eigenvector. Since the eigenvectors of each factor graph are trigonometric functions the inner products can be evaluated using a Discrete Fourier Transform (DFT) [70]. The DFT of a signal $x = [x_1, x_2, ...x_n]$ is defined:

$$\hat{x}_k = \sum_{j=1}^{V} \exp\left(-i\frac{2\pi}{n}(k-1)(j-1)\right)x_j. \tag{3.59}$$

Let $\mathcal{F}$ denote the DFT so $\hat{x} = \mathcal{F}(x)$. Note that the DFT is not equivalent to the inner

product with the eigenvectors of a line or cycle (see Equation (3.53) and Equation (3.54)). In order to use the DFT to move into the eigenbasis we need to work out a scaling of the DFT that will produce the desired coefficients.

First consider a factor graph that is a line of V nodes. Then, to transform a vector $y = [y_1, ...y_V] \in \mathbb{R}^V$ into the eigenbasis of the node Laplacian we need to evaluate the inner products:

$$\hat{y}_1 = V^{-1/2} \sum_{j=1}^{V} y_j = 2^{-1/2} \sum_{j=1}^{V} \left(\frac{V}{2}\right)^{-1/2} \cos\left(\frac{\pi(1-1)}{V}(j-0.5)\right) y_j$$
$$\hat{y}_k = \sum_{j=1}^{V} \left(\frac{V}{2}\right)^{-1/2} \cos\left(\frac{\pi(k-1)}{V}(j-0.5)\right) y_j.$$

Therefore we need to be able to evaluate inner products with the vectors whose entries are specified by the trigonometric function $\cos\left(\frac{\pi}{V}(k-1)(j-0.5)\right)$. These inner products can be evaluated by scaling the DFT of a modified version of the signal y. Let $x = [y; \mathbf{0}]$ where $\mathbf{0}$ is the $V \times 1$ vector of all zeros. Then:

$$\begin{aligned}
\hat{x}_k &= \sum_{j=1}^{2V} \exp\left(-i\frac{2\pi}{2V}(k-1)(j-1)\right) x_j \\
&= \sum_{j=1}^{V} \exp\left(-i\frac{\pi(k-1)}{V}(j-1)\right) y_j \\
&= \exp\left(i\frac{\pi}{2V}(k-1)\right) \sum_{j=1}^{V} \exp\left(-i\frac{\pi(k-1)}{V}(j-0.5)\right) y_j.
\end{aligned}$$

Therefore:

$$\begin{aligned}\exp\left(-i\frac{\pi(k-1)}{2V}\right)\hat{x}_k &= \sum_{j=1}^{V}\exp\left(-i\frac{\pi(k-1)}{V}(j-0.5)\right)y_j \\ &= \sum_{j=1}^{V}\cos\left(\frac{\pi(k-1)}{V}(j-0.5)\right)y_j \\ &\quad - i\sum_{j=1}^{V}\sin\left(\frac{\pi(k-1)}{V}(j-0.5)\right)y_j.\end{aligned}$$

Since all the entries of y are real:

$$\sum_{j=1}^{V}\cos\left(\frac{\pi(k-1)}{V}(j-0.5)\right)y_j = \text{Real}\left(\exp\left(-i\frac{\pi(k-1)}{2V}\right)\hat{x}_k\right).$$

The inner product on the left hand side is proportional to the inner product of y with the k^{th} eigenvector. Therefore, if $\mathcal{G}_j$ is a line with V_j nodes the expansion of $y \in \mathbb{R}^{V_j}$ onto the eigenbasis is given by:

$$\begin{aligned}\hat{x} &= \mathcal{F}([y;\mathbf{0}]) \\ \hat{y}_k &= \left\{\begin{array}{l} V_j^{-1/2}\text{Real}\left(\hat{x}_1\right) \text{ if } k=1 \\ \left(\frac{V_j}{2}\right)^{-1/2}\text{Real}\left(e^{-i\frac{\pi(k-1)}{2V}}\hat{x}_k\right) \text{ if } k>1\end{array}\right\}.\end{aligned} \tag{3.60}$$

The DFT can be performed efficiently using the fast Fourier transform (FFT). The cost of the FFT is order $n\log(n)$ where n is the length of the sequence transformed [70]. Therefore the computation cost of Equation (3.60) is order $2V\log(V)$, while the computation cost of performing each inner product directly is order V^2.

Essentially the same method can be used if the factor graph is a cycle instead of a line. To expand into the eigenbasis associated with a cycle of length V we need inner products

with $\cos(\pi(k-1)/V(j-0.5))$ and $\sin(\pi k/V(j-0.5))$. The former inner products can be computed in the same fashion as on the line. To compute the inner products with $\sin$ take the imaginary part of the shifted DFT instead of the real part:

$$\sum_{j=1}^{V} \sin\left(\frac{\pi(k-1)}{V}(j-0.5)\right) y_j = -\text{Im}\left(\exp\left(-i\frac{\pi(k-1)}{2V}\right)\hat{x}_k\right).$$

Then:

$$\sum_{j=1}^{V} \sin\left(\frac{\pi k}{V}(j-0.5)\right) y_j = -\text{Im}\left(\exp\left(-i\frac{\pi k}{2V}\right)\hat{x}_{k+1}\right).$$

Therefore, if $\mathcal{G}_j$ is a cycle with V_j nodes, then the expansion of $y \in \mathbb{R}^{V_j}$ onto the eigenbasis is given by:

$$\begin{aligned}
&\hat{x} = \mathcal{F}([y; \mathbf{0}]) \\
&\hat{y}_k = \left\{ \begin{array}{l}
V_j^{-1/2}\text{Real}\left(\hat{x}_1\right) \text{ if } k = 1 \\
-\left(\frac{V_j}{2}\right)^{-1/2} \text{Im}\left(e^{-i\frac{\pi k}{2V}}\hat{x}_{k+1}\right) \text{ if } k \in [2, V-1] \text{ and even} \\
\left(\frac{V_j}{2}\right)^{-1/2} \text{Real}\left(e^{-i\frac{\pi(k-1)}{2V}}\hat{x}_k\right) \text{ if } k \in [2, V] \text{ and odd} \\
-V_j^{-1/2}\text{Im}\left(e^{-i\frac{\pi k}{2V}}\hat{x}_{k+1}\right) \text{ if } k = V \text{ and } V \text{ is even}
\end{array} \right\}.
\end{aligned} \tag{3.61}$$

As when $\mathcal{G}_j$ is a line, this method runs in order $V_j \log(V_j)$ rather than order V_j^2 if the DFT is implemented with an FFT.

Therefore, if $\mathcal{G}$ is the product of a sequence of line graphs and loops then solution to the discrete Poisson equation $L_{\mathcal{V}}^2$ on the eigenbasis, $\hat{\phi}$ can be solved as follows:

DFT Algorithm for Solving the Discrete Poisson Equation in the Eigenbasis of a Lattice:

1. Compute the divergence of the edge flow at every node in the product graph. Use the multi-index $j = (j_1, j_2, ...j_d)$ to represent the nodes of the product graph. Let d_j equal to divergence at node j.

2. Set $\hat{d}_j = d_j$ for all j.

3. Loop over the factor graphs $\mathcal{G}_n$ from $n = 1$ to d

   (a) Form a vector $y(h)$ for each possible state of the product of the remaining factor graphs $h = (j_1, j_2, ...j_{n-1}, j_{n+1}, ..., d)$ where $y(h)_l$ is $\hat{d}_j$ at state $j = (j_1, ...j_{n-1}, l, j_{n+1}, ..., d)$ in the full product.

   (b) Let $x(h) = [y(h); \mathbf{0}]$ and use an FFT to compute $\hat{x}(h) = \mathcal{F}(x(h))$.

   (c) If $\mathcal{G}_j$ is a line use Equation (3.60) to recover $\hat{y}(h)$. If $\mathcal{G}_j$ is a cycle use Equation (3.61) to recover $\hat{y}(h)$.

   (d) Set $\hat{d}_j$ equal to $\hat{y}_j$ for all $j = (j_1, j_2, ...j_{n-1}, l, j_{n+1}, ..., d)$

4. For each $\hat{d}_j$ compute the eigenvalues $\lambda_{j_1}, \lambda_{j_2}, ..., \lambda_{j_d}$ where λ_{j_n} is defined by either Equation (3.53) or Equation (3.54) if the n^{th} factor graph is a lien or a cycle (respectively).

5. Then compute $\hat{\phi}_j = \hat{d}_j \sum_{n=1}^{d} \lambda_{j_n}$.

Then, to recover the solution we repeat step 3., only replacing the divergence with $\hat{\phi}$ and scaling an iFFT instead of an FFT to recover the product with the matrix of eigenvectors instead of the matrix of eigenvectors transpose. The necessary scaling can be calculated in the same way we derived Equation (3.60) and Equation (3.61).

The method outlined above is essentially a multi-dimensional FFT. The FFT is per-

formed one factor graph at a time to transform the divergence of f into the eigenbasis of the node Laplacian. Then the transformed divergence is scaled by the eigenvalues of the node Laplacian (sums of eigenvalues of the factor graph), and a multidimensional iFFT is used to recover the solution. While more involved than applying the matrix of eigenvectors directly (which can be formed by the repeated Kronecker product of the matrix of eigenvectors of each product graph), this method is much more efficient and conceptually consistent with the solution to the Poisson equation in the continuum, and to its discrete approximation (cf. [73]).

The cost of each of the d FFT and iFFT steps is $V_j \log(V_j)$ [70], and for each dimension d we perform $\prod_{n \neq j} V_n = V/V_j$ transforms. Thus the cost for each dimension/factor graph is $(V/V_j)V_j \log(V_j) = V \log(V_j)$. Then, summing over the d dimensions, the total cost is $\sum_j V \log(V_j) = V \log(\prod_j V_j) = V \log(V)$, which matches the standard runtime of a multidimensional FFT [70]. Therefore the overall computational cost of the FFT based approach to solving the discrete Poisson equation is $V \log(V) = (\prod_{j=1}^{d} V_j) \sum_{j=1}^{d} \log(V_j)$. In contrast, the computational cost for solving the discrete Poisson equation by performing the transform into and out of the eigenbasis directly (with a matrix product) would be $V^2 = (\prod_{j=1}^{d} V_j)^2$. Consequently, if any of the product graphs are large, or d is large, then the FFT based method is much more efficient than direct application of the matrix of eigenvectors.

3.5 Numerical Methods for General Networks

This section introduces generic methods for constructing the operators and performing the HHD on an arbitrary network.

Let $\mathcal{G}$ be a finite connected network with V vertices and E edges. Networks are stored

efficiently via an adjacency structure [32]. An adjacency structure is a set of lists, one for each node in the network. The list corresponding to node j is the set of all nodes who are connected to j by an edge. That is, the adjacency list associated with node j is the set of neighbors of j. An adjacency structure can be computed from an adjacency matrix by finding all the nonzero entries in the row of an adjacency matrix, and is a sparse representation of the adjacency matrix of a graph [32].

Assume that the graph is stored in an adjacency structure, but no more information on the topology is provided. How can the operators be constructed using only the adjacency structure?

In order to construct the operators we need an ordering on the vertices, an ordering on the edges and reference orientation for each edge, and an oriented cycle basis. Usually the vertices in the adjacency structure are referenced by an index, so it is reasonable to assume that the vertices are ordered a priori. Therefore our first task is to introduce an ordering on the edges.

By convention let each edge point from a lower indexed node to a higher indexed node. Then for edge k the $i(k) < j(k)$. Then it is natural to order the edges in lexicographic order. This can be accomplished as follows. Start with the adjacency list for the first vertex. Order the neighbors of first vertex in increasing order. Then, for each neighbor of the first vertex introduce an edge, and number the edges according to which neighbor of the first vertex they point to. For example, if node 1 neighbors nodes 3, 5, and 11 then edges 1, 2, and 3, are the edges from 1 to 3, 1 to 5, and 1 to 11.

To store the endpoints of each edge introduce an edge to endpoint mapping $M \in \mathbb{Z}^{E,2}$, where $M(k,1) = i(k)$ is the start of edge k and $M(k,2) = j(k)$ is the end of edge k. Each time we add an edge fill in the corresponding row of the edge to endpoint mapping. So, in the example, the first three rows of M are $[1,3], [1,5], [1,11]$. After all of the edges leaving

the first node have been indexed and added to M move on to the second node. Order the neighbors of the second node in increasing order. Then add an edge for each neighbor of the second node that is not the first node. Add the matching rows to to M and repeat the process for the third node. For each node we add an edge for every neighbor who has a higher index than the node considered. This is equivalent to working across the rows of the adjacency matrix and storing a new edge for each nonzero entry above the diagonal.

Once the process is complete the edges will be ordered lexicographically by their endpoints, and the matrix M will store the endpoints of every edge. Given M, a sparse representation of the gradient can be easily constructed. Starting with the first column of M, add a negative one to each row of the gradient in entry $k, i(k) = k, M(k, 1)$. Then add a one to each row of the gradient in entry $k, j(k) = k, M(k, 2)$. This can be easily implemented using a sparse matrix command. For example, in Matlab, the command: G = sparse((1:E),M(:,2),1,E,V) - sparse((1:E),M(:,1),1,E,V) will construct the gradient directly from the edge to endpoint mapping.

The node Laplacian can then be computed either by computing $G^{\mathsf{T}}G$, or by computing the degree of each node, and subtracting the adjacency matrix from the degree matrix.

Once the gradient and Laplacian are computed it is easy to compute the scalar potential, conservative, and rotational edge flows. If the network is small then the discrete Poisson equation $L_{\mathcal{V}}^2\phi = -G^{\mathsf{T}}f$ could be solved directly, or the projector onto the conservative subspace could be found via a QR decomposition of the gradient (see Section 2.3). Alternatively, the least squares problem $\text{argmin}_{u\in\mathbb{R}^V}\{||Gu + f||\}$ could be solved with an iterative least squares solver. Since iterative solvers only require the forward application of the gradient all that it required for the iterative solvers is an implicit representation of the gradient and divergence. These are easily implemented from the edge to endpoint mapping. Loop over each edge k, find the endpoints from $M(k, :)$, and compute $u_{j(k)} - u_{i(k)}$. To

implement the divergence implicitly loop over each edge k, find the endpoints of the edge from $M(k,:)$, and add f_k to the divergence of $i(k)$ and subtract it from the divergence of $j(k)$.

Once ϕ is found, the conservative field is easily computed using $f_{\text{con}} = -G\phi$. The rotational field follows immediately, $f_{\text{rot}} = f - f_{\text{con}}$. Therefore the scalar potential, conservative, and rotational fields can all be easily found provided only the information contained in an adjacency structure.

Note that, the flow f may be provided in an antisymmetric matrix with i, j entries $f_{i,j}$, in which case the vector $f \in \mathbb{R}^E$ corresponding to the edge ordering should be constructed one entry at a time as the edges are indexed. If f is provided as a vector to start then the edges must be indexed and oriented a priori so no work is needed to construct the edge ordering and orientation. Alternatively the edge flow f may be a function of the endpoints, so that given $i(k), j(k)$ the flow f_k can be computed. This would be the case if the endpoints represent states of a system, and the edge flow represents transition rates between states that depend on the states at either end of each edge (see Chapter 6).

All that remains is to construct the curl and recover the rotational potential. This is not as easy since the curl requires a cycle basis, and it requires work to construct a cycle basis. Moreover, since cycle bases are rarely unique, but often differ in quality, we will have to tailor our method of cycle basis construction to arrive at a desired cycle basis. The rest of this section is devoted to the problem of finding a desired cycle basis, and efficiently constructing the associated curl.

3.5.1 Desired Properties of Cycle Basis

What do we want from a cycle basis? The answer depends in part on interpretation and in part on numerics.

If the only goal were to recover θ then the best cycle basis to choose would be a fundamental cycle basis. The cycles in a fundamental cycle basis each correspond to a chord, and no two cycles ever cross the same chord. As an immediate consequence, the rotational flow across a chord is the value of the rotational potential θ on that chord, so once the rotational flow is separated from the total flow recovering the rotational potential is trivial. The principal downside to this approach is that it tells us nothing new. The rotational potential is the same as the rotational flow on the chords, so provides no new insight into the circulation of the edge flow.

The other principal downside to working with fundamental cycle bases is that they typically consist of large sequences of nested loops (see Figure 3.17). This nesting occurs because the only way to complete any loop is to trace back to a point where the tree branches. Thus two loops sharing the same branches tend to nest inside of each other. As a result, whole sequences of loops are larger than necessary, and overlap. The resulting cycle basis is unbalanced, with edges in the tree appearing in many loops, while chords only appear in one loop. If the tree is generated using a search procedure that radiates out from a central node then the cycles in the fundamental cycle basis will also vary in size, with chords far removed from the central node forming large loops, while chords close to the central node form small loops. Worse, the pairs of large nested loops often only differ by a few edges, so much of the loop representation is redundant.

If the graph is planar then the obvious objective is to recover a planar cycle basis. A planar cycle basis contains all but one of the faces of the planar graph [45]. The cycles in a planar cycle basis should be oriented so that any pair of cycles that share an edge cross it in opposite directions. If a planar cycle basis is used then no edge is included in more than two basis cycles, and if it is possible to find a cycle basis with this property then the graph is necessarily planar [35]]. A fundamental cycle basis usually won't match the faces

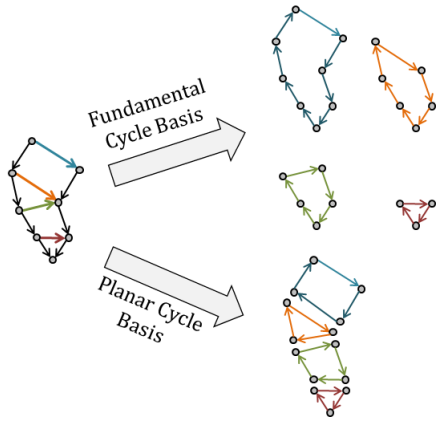

Figure 3.17: A fundamental cycle basis and a planar cycle basis generated by the chords connecting two branches of a spanning tree. Note that the fundamental basis contains a sequence of large nested loops instead of a tiling of small loops. Notice that all four loops in the fundamental basis share the same two edges at the branching point, whereas in the planar basis no edge borders more than two loops.

a planar graph, so will not usually recover a planar basis.

One way to optimize of a cycle basis is to reduce the number of times each edge appears in the basis. Reducing the number of occurrences of each edge reduces redundancy in the cycles, so ought to lead to a better conditioned face Laplacian. Moreover, if the graph is planar then minimizing the max over all edges of the number of times an edge appears in the cycle basis guarantees that the cycle basis is planar.

An obvious alternative is to attempt to find a cycle basis with small cycles. The smaller the cycles the more localized the effects of the rotational potential on a given loop. Minimizing the size of the cycles also reduces the number of times cycles overlap since the total number of nonzero entries in the curl is minimized. Thus minimizing cycle length promotes independent cycles and avoids intertwined cycles. Cycle bases that are sparse in this sense are widely preferred in applications [33].

The length of a cycle basis $|\mathcal{C}|$ is the sum of the length of each cycle in the basis, $\sum_j |\mathcal{C}_j|$. A minimal cycle basis is a cycle basis that minimizes $|\mathcal{C}|$ [33]. Minimal cycle

bases (MCB) are widely sought in applications (cf. [65]), and there exist a number of algorithms for finding an MCB, or approximation to an MCB, both in generic graphs and for special classes of graphs [42, 81, 82].

Note that $|\mathcal{C}|$ is the sum over the rows, of the number of nonzero entries in each row. This equals the sum over the columns of the number of nonzero entries in each column, therefore the average cycle length $|\mathcal{C}|/E$ equals the average number of of times any edge appears in the basis. Since the number of edges E is fixed, a minimal cycle basis also minimizes the reuse of edges.

Minimal cycle bases are an example of optimal cycle bases. An optimal cycle basis is a basis that optimizes some criteria, for example, the number of reoccurences of an edge, or the total length of all cycles in the basis. Considerable effort has been devoted to developing algorithms for optimizing cycle bases. Finding an optimal cycle basis can be a difficult combinatorial problem, especially for large graphs, and while some efficient algorithms exist it is often expensive to find an optimal cycle basis if the graph is large. A useful review of algorithms for finding minimal cycle bases is available in [33]. For some classes of cycle basis it is possible to find a minimal cycle basis in polynomial time, while for others it is APX hard. However, even when it is possible to find polynomial time algorithms they are often expensive [33, 42].

Kavitha presents a deterministic polynomial time algorithm that runs in $\mathcal{O}(E^2\frac{V}{\log(V)} + EV^2)$ and notes that most minimal cycle bases algorithms have space costs $\mathcal{O}(E^2)$ so are not applicable to large graphs [33]. Horton provided a greedy polynomial time algorithm that works by first forming a set of $\mathcal{O}(EV)$ circuits, adding cycles to a list of possible basis cycles in order from shortest to longest, and reducing the set via Gaussian elimination modulo 2 to check independence at each step. While straightforward, this algorithm runs in $\mathcal{O}(E^3V)$ time [42]. The expense of MCB algorithms motivates approximate MCB

algorithms. For example, Horton presents a modified version of his algorithm which runs in $\mathcal{O}(EV^2)$ time [42].

Alternatively, we may desire a cycle basis in which the cycles are as independent as possible. That is, a cycle basis with as little redundancy in the basis cycles as possible. This requires that there is no rotational potential $v \in \mathbb{R}^L$ such that $||v||_2$ is large, but $||C^\intercal v||_2$ is small. The ratio of the norm of the flow to the norm of the potential is the square root of the Rayleigh quotient of $C^\intercal$ and v [51]. The square root is monotonically increasing, so minimizing the Rayleigh quotient is the same as minimizing the root of the Rayleigh quotient. Thus, a natural measure of the redundancy of the cycle basis with respect to a potential v is:

$$R(CC^\intercal, v) = \frac{||C^\intercal v||_2^2}{||v||_2^2} = \frac{v^\intercal CC^\intercal v}{v^\intercal v} = \frac{v^\intercal L_{\mathcal{C}}^2 v}{v^\intercal v}. \tag{3.62}$$

Then the associated measure of redundancy in the basis is $\min_{v\in\mathbb{R}^L}\{R(CC^\intercal, v)\}$. The numerical range of $CC^\intercal$ is the range of possible values of the Rayleigh quotient. Since $CC^\intercal$ is a real symmetric graph it is unitarily diagonalizable, so the numerical range is the convex hull of the eigenvalues of $CC^\intercal$ [51], which equal its singular values since the matrix is real symmetric and positive semi-definite. Therefore the minimum value of the Rayleigh quotient is the smallest singular value of the face Laplacian:

$$\min_{v\in\mathbb{R}^L}\{R(CC^\intercal, v)\} = \sigma_L(CC^\intercal) = \min\{\sigma(L_{\mathcal{C}}^2)\}. \tag{3.63}$$

An optimal cycle basis with respect to Equation (3.63) solves the max-min problem:

$$\max_{\mathcal{C}\in\mathcal{B}}\{\min\{\sigma(L_{\mathcal{C}}^2)\}\}$$

where $\mathcal{B}$ is the set of all cycle bases of the graph $\mathcal{G}$. A partial ordering of the set of cycle

bases can be introduced setting $\mathcal{C} > \hat{\mathcal{C}}$ if the smallest singular value of $L^2_{\mathcal{C}}$ is greater than the smallest singular value of $L^2_{\hat{\mathcal{C}}}$, and with equality if the two singular values are equal. Alternatively, if the two smallest singular values are different, then compare the second smallest singular values of the face Laplacians. Continuing in this manner $\mathcal{C} > \hat{\mathcal{C}}$ if the last singular value of $L^2_{\mathcal{C}}$ that is different from the corresponding singular value of $L^2_{\hat{\mathcal{C}}}$ is larger than the corresponding singular value of $L^2_{\hat{\mathcal{C}}}$.

We may also want to optimize the cycle basis to promote stability in the inverse problem $CC^\intercal\theta = Cf$. Then a natural objective function is the condition number κ of the face Laplacian $L^2_{\mathcal{C}}$ (ratio of largest and smallest singular values). Since the smallest value of the Rayleigh quotient is the smallest singular value of the Laplacian, and the largest value of the Rayleigh quotient is the largest singular value [51], the condition number of the face Laplacian can be bounded using the Rayleigh quotient. We can bound the condition number of the Laplacian from below by finding a v such that $\frac{||C^\intercal v||^2}{||v||^2}$ is large, which gives a lower bound on the largest eigenvalue, and a v such that $\frac{||C^\intercal v||^2}{||v||^2}$ is small, which gives an upper bound on the smallest eigenvalue.

Consider a cycle basis $\mathcal{C}$ with corresponding curl C. Each loop in the cycle basis corresponds to a row in C. Denote the length of the j^{th} basis loop $|C_j|$. Let $v = e_j$ where e_j is the j^{th} column of the identity. Then $C^\intercal e_j$ is equivalent to the j^{th} row of the curl. This row has precisely $|C_j|$ nonzero entries all equal to one or negative one, so $||e_j^\intercal C^\intercal C e_j|| = |C_j|$. It follows that the largest eigenvalue of the face Laplacian must be greater or equal to the length of the largest loop in the basis:

$$\sigma_1(\mathcal{C}) = \max\{\sigma(L^2_{\mathcal{C}})\} \geq \max_{C \in \mathcal{C}}\{|C|\}. \tag{3.64}$$

By the same logic:

$$\sigma_F(\mathcal{C}) = \min\{\sigma(L_{\mathcal{C}}^2)\} \leq \min_{C\in\mathcal{C}}\{|C|\}. \tag{3.65}$$

This means that the condition number of C is bounded from below by the ratio of the perimeter of the largest loop in the basis to the perimeter of the smallest loop in the basis:

$$\kappa(\mathcal{C}) \geq \frac{\max_{C\in\mathcal{C}}\{|C|\}}{\min_{C\in\mathcal{C}}\{|C|\}}. \tag{3.66}$$

This means that cycle bases with loops of wildly different sizes will tend to be ill-conditioned, so, when possible, it is good to look for a basis with relatively uniformly sized loops. Notice that if the loops are generated by a breadth first search then the maximum perimeter loop will usually scale in $\log(V)$, so using a breadth first search will ensure that, even for large networks, the ratio of the longest loop to the shortest loop is not overly large.

This bound was derived by only considering w of the form e_j. The problem with nested loops is not that one is much larger than the other, it is that both are large, but their combination is small. This means that their corresponding rows in the curl are identical in most of their entries, and only different in a small subset of their nonzero entries. In that case their difference is much smaller than their magnitudes, so they are close to linearly dependent.

Consider all pairs of adjacent basis loops j, k. Then for each pair pick $w = e_j \pm e_k$ where plus is used if the loops run in opposite directions on their shared path, and minus is used if they run in the same direction. Then $||C^\intercal w|| = |C_j \triangle C_k|$ where $|C_j \triangle C_k|$ denotes the perimeter of the symmetric difference of the two loops (their linear combination). Then:

$$\sigma_F(\mathcal{C}) = \min\{\sigma(L_{\mathcal{C}}^2)\} \ \leq \frac{1}{2}\min_{C_j,C_k\in\mathcal{C}}\{|C_j \triangle C_k|\}. \tag{3.67}$$

where the factor of 2 in the denominator comes from the norm of the vector $v = e_j \pm e_k$.

Therefore the condition number of the face Laplacian is bound from below by:

$$\kappa(\mathcal{C}) \geq \frac{\max_{\mathcal{C}_j \in \mathcal{C}}\{|\mathcal{C}_j|\}}{\min\{\min_{\mathcal{C}_j \in \mathcal{C}}\{|\mathcal{C}_j|\}, \min_{\mathcal{C}_j, \mathcal{C}_k \in \mathcal{C}} \left\{\frac{|\mathcal{C}_j \triangle \mathcal{C}_k|}{2}\right\}\}}. \tag{3.68}$$

Thus the condition number is not only greater than the ratio of the largest loop to the smallest loop, but is also greater than the ratio of the largest loop to the smallest possible loop formed by combinations of pairs of loops, all multiplied by 2. This result reveals the problem with large nested loops. Large loops ensure that the numerator is large, while the small difference between the loops ensures the denominator is small. It follows that we will generally want a cycle basis with loops of approximately uniform size, and that are as close to pairwise independent as possible.

The fact that fundamental cycle bases reuse the edges in the spanning tree many times and contain nested loops that are close to redundant suggests that the linear system used to recover θ should be unstable for large graphs. The linear system is not unstable because, as noted before, the structure of fundamental cycle bases makes recovering θ from f easy once f_{rot} is known. The rotational field can be recovered directly from the residual, $G\phi + f$ when solving for ϕ, thus we can find f_{rot} using operators that are entirely cycle basis independent. Recovering θ from f_{rot} is easy on a fundamental cycle basis since each loop in the basis crosses an edge that is not a part of any other loop. In fact, if $\mathcal{C}$ is an arbitrary cycle basis, and $\mathcal{C}_j$ includes an edge, k, which is not included in any other loop in the basis then $\theta_j = \pm f_{\text{rot}k}$. This observation is useful since it points towards a broader class of cycle bases which are more general than fundamental cycle bases, but also allow for nearly trivial calculation of θ from f_{rot}.

We say that a collection of cycles $\mathcal{C}$ has a boundary if there is a cycle in $\mathcal{C}$ that includes

an edge that is not included in any other cycle. We say that a cycle $\mathcal{C}_j$ is in the boundary of the collection of cycles if it crosses an edge not crossed by any other cycle in the set. Consider a set of cycles such that (i) removing any boundary cycle from the set produces a new set of cycles with a boundary, and (ii) if cycles on the boundary are removed iteratively, then the remaining cycle set always has a boundary. Then this set of cycles is a set that still has a boundary after a sequence of boundary cycles are removed. Moreover, there must exist a permutation σ such that:

$$\mathcal{C}_{\sigma(i)} \setminus (\cup_{j=1}^{i-1} \mathcal{C}_{\sigma(j)}) \neq \emptyset.$$

That is, there is an ordering of the cycles so that, cycle i includes at least one edge not included in any earlier cycle. A set of L cycles with this property is a weakly fundamental cycle basis [33].

Lemma 13 (Weakly Fundamental Cycle Basis)**.** *A set of L cycles is a cycle basis if there exists a ordering of the cycles σ such that:*

$$\mathcal{C}_{\sigma(i)} \setminus (\cup_{j=1}^{i-1} \mathcal{C}_{\sigma(j)}) \neq \emptyset. \tag{3.69}$$

Then the linear system $C^\intercal \theta = f_{rot}$ can be solved by the following iteration:

1. *Initialize: $j = 0$, $\hat{f}(j) = f_{rot}$, $\hat{\mathcal{C}}(j) = \mathcal{C}$*
2. *Iterate from $j = 0$ to $j = L - 1$:*
   (a) *Find a cycle, $\mathcal{C}_h$, in the boundary of $\hat{\mathcal{C}}(j)$. Let k denote the edge in the boundary cycle crossed by no other cycle in $\hat{\mathcal{C}}(j)$. Set $\theta(h) = \hat{f}(j)_k$ if cycle $\mathcal{C}_h$ crosses edge k in its forward direction, and set $\theta(h) = -\hat{f}(j)_k$ if cycle $\mathcal{C}_h$ crosses edge*

k in its backward direction.

(b) Remove the boundary cycle: $\hat{f}(j+1) = \hat{f}(j) - C^{\intercal}(\mathcal{C}_h)\theta_h$, $\hat{\mathcal{C}}(j+1) = \hat{\mathcal{C}}(j) \setminus \mathcal{C}_h$.

Proof. Let $\mathcal{C}$ be a set of L cycles that satisfies Equation (3.69). Since the set contains L cycles it is a basis if the cycles are linearly independent. Since the set of cycles satisfies Equation (3.69) it is possible to order the cycles of $\mathcal{C}$ so that $\mathcal{C}_1$ is a boundary cycle, and once $\mathcal{C}_1$ through $\mathcal{C}_j$ are removed $\mathcal{C}_{j+1}$ is a boundary cycle. It is clear that $\mathcal{C}_1$ is independent of $\{\mathcal{C}_2, ..., \mathcal{C}_L\}$ since it includes an edge that none of the other cycles include. Thus there is no way to combine the cycles in $\{\mathcal{C}_2, ..., \mathcal{C}_L\}$ to produce $\mathcal{C}_1$, so the only way that $\mathcal{C}$ could be linearly dependent is if the set $\{\mathcal{C}_2, ..., \mathcal{C}_L\}$ are linearly dependent. But the set $\{\mathcal{C}_2, ..., \mathcal{C}_L\}$ also has a boundary $\mathcal{C}_2$ is a boundary cycle for the reduced set. Therefore $\mathcal{C}_2$ is necessarily independent of $\{\mathcal{C}_3, ..., \mathcal{C}_L\}$ since it includes an edge not included by any other cycles in the reduced set. Repeating this argument inductively shows that any simple set of cycles is a set of independent cycles, so if $\mathcal{C}$ is a simple set of L cycles it must be a cycle basis.

To prove that θ satisfying $C^{\intercal}\theta = f_{\text{rot}}$ is recovered by the iterative procedure note that, since $\mathcal{C}_h$ is a boundary cycle of $\mathcal{C}$, then it includes an edge, k, not included in any other cycle. Therefore $\theta_h = \pm f_{\text{rot}\,k}$ where the sign depends on which direction the cycle crosses the edge. Thus one entry of θ can be recovered directly. Since the systems of equations $C^{\intercal}\theta = f_{\text{rot}}$ is linear we can subtract $C^{\intercal}(\theta_h e_h)$ from both sides and then remove the column corresponding to cycle h from the systems. Then we are left with $C^{\intercal}(\mathcal{C} \setminus \mathcal{C}_h)$ on the left hand side and $f_{\text{rot}} - C^{\intercal}(\mathcal{C}_h)\theta_h$ on the right hand side. Now the systems of equations has one fewer column, and one fewer unknown, but is of the same form as before since $\mathcal{C} \setminus \mathcal{C}_h$ still satisfies satisfies Equation (3.69). Thus the process can be repeated iteratively, solving for θ on a boundary cycle, subtracting the corresponding cyclic flow from the right hand side, and removing the boundary cycle from the set. Since there are finitely many cycles in

the set this process ends after L steps, thereby recovering θ.

□

Note that, if the loops are ordered so that $\mathcal{C}_j$ is a boundary loop for $\{\mathcal{C}_{j+1}, ..., \mathcal{C}_L\}$ and the edges are ordered so that f_j is an edge in $\mathcal{C}_j$ crossed by no other cycle in $\{\mathcal{C}_{j+1}, ..., \mathcal{C}_L\}$, then $C^\intercal$ is lower triangular with diagonal entries equal to ± 1, and the iterative method proposed is the same as back-substitution. It follows that this method is backwards stable [51].

Weakly fundamental cycle bases are more general than fundamental cycle bases since all fundamental cycle bases satisfy Equation (3.69), but not all weakly fundamental cycle bases are fundamental. Weakly fundamental cycle bases include much more intuitive cycle bases than the class of fundamental cycle bases. For example:

Lemma 14 (Planar Bases are Weakly Fundamental)**.** *If $\mathcal{G}$ is a finite connected planar graph then any planar cycle basis of $\mathcal{G}$ is a weakly fundamental cycle basis.*

Proof. If $\mathcal{G}$ is a finite connected planar graph then any set of L faces of the graph is a planar cycle basis. The graph includes a total of $L + 1$ faces, L on the interior and one external face. Embed the graph so that the face excluded from the basis is the external face. Then all of the edges bordering the external face are on the boundary of the graph and neighbor at most one face from the interior (no edge in a planar graph borders more than two faces). Since the interior faces are cycles in the cycle basis there are cycles in the cycle basis that border edges included in no other cycle from the cycle basis. These are the cycles that border the boundary of the embedded graph, hence the choice to call these cycles boundary cycles. Consider the subgraph of $\mathcal{G}$ with only the nodes and edges that appear in the set of cycle basis. If a boundary cycle is removed then all nodes and edges that are included in that boundary cycle and no other are pruned from the subgraph. The

pruned subgraph is still planar, so it still includes boundary cycles. Thus the set of cycles left over after removing a sequence of boundary cycles still has a simple boundary since the corresponding subgraph is still planar. Since the graph is finite the number of cycles in the cycle basis is finite, so after L boundary cycles are removed no cycles are left, thus the set of L cycles was a weakly fundamental cycle basis. □

In contrast, if $\mathcal{G}$ is a planar graph with some faces that do not border the boundary, then no matter which face is chosen as the exterior face, no planar basis of $\mathcal{G}$ is ever a fundamental cycle basis since all of the edges of the interior faces that do not border the boundary border one other cycle in the cycle basis, and in a fundamental basis all cycles include an edge not included in any other basis cycle.

These considerations set the stage for the rest of the chapter. In Section 3.5.2 we introduce a simple search procedure for constructing a fundamental cycle basis. In Section 3.5.3 we introduce a greedy search procedure that produces a weakly fundamental cycle basis, and that can be designed to promote small cycles and to reduce the number of repeated edge crossings. This search procedure is shown to be efficient, and is tested on random graphs with up to a million vertices. To conclude we discuss an alternative search procedure for planar graphs Section 3.5.4. The search procedure is designed to recover a planar basis, and is based on iteratively partitioning the planar graph into smaller graphs.

3.5.2 Constructing a Fundamental Cycle Basis

A fundamental cycle basis is defined by a spanning tree $\mathcal{T}$. The first task when constructing a fundamental cycle basis is to construct the tree $\mathcal{T}$. A spanning tree can be constructed with a search procedure. Both depth first search and breadth first search procedures can be used to build spanning trees out from a central node. Our procedure is based on breadth

first search since it produces bushier trees, and as a result, shorter cycles. Some authors use depth first searches instead (cf. [42]). We use a breadth first search so that, by the "small world effect" [83], the loops in our fundamental cycle basis are not excessively large.

Start by computing the degree of each node. Reorder the nodes in decreasing order of degree. Then initialize a breadth first search from the node with maximal degree. As the search progresses we store: the edges in the tree, the chords, the parent edge of each node, and, depending on the implementation, the ancestral edges each node. The parent edge of a node in the tree is the edge leaving that node on the path back to the root of the tree (node used to initialize the search). The list of ancestral edges is the collection of all edges in the path from a node back to the root. The cost of storing the ancestral edges depends on the lengths of these paths. By using a breadth first search instead of a depth first search we guarantee that no list of ancestral edges includes more edges than the diameter of the graph, and no list of ancestral edges is longer than the longest distance from any node to the root of the tree. For random graphs the diameter is almost always logarithmic in the number of nodes with base equal to the average degree if the graph is sparse [84], and average distance between randomly chosen nodes scales in the $\log$ base d of the number of vertices where d is the average degree[5] of the nodes in the network [85], so the storage cost of storing the ancestral edges for each node in the tree is expected to scale with $V \log_d(V)$. The more positively skewed the degree distribution the larger the base of the logarithm, thus the smaller the average distance between nodes, and the more clustered the graph the larger the average distance between randomly chosen nodes. The advantage of storing the ancestral edges is that it streamlines construction of the basis loops. If only the parent edges are stored then we need to search backwards from the endpoints of each chord to find each basis loop.

[5]To be precise, d should be a weighted sum of squares of the expected degrees.

At each stage of the search we have a partial tree. At stage n the leaves of the tree are all nodes a distance n from the initial node, where distance the distance between two nodes is the length of the shortest path between them. Loop in order from the leaf with largest degree to the leaf with smallest degree. For each leaf find all the neighbors of that leaf that have not yet been added to the tree. Add these neighbors to the tree and add the edges from the leaf to the list of edges in the tree. Set the parent edge of each new neighbor to the edge used to find it from the leaf. Set the ancestral edges for any new neighbor to the ancestral edges of the leaf plus the parent edge of the new neighbor. If a neighbor is found that has already been added to the tree then the corresponding edge is a chord. Add the chord to the list of chords if it has not been found before. This can be done automatically without searching the list of chords by only adding a chord if the neighbor at then end of the chord has not yet been searched from. Repeat this process for all the nodes a distance n from the initial node, then iterate until all nodes and edges have been found.

Once the search is complete we have a spanning tree $\mathcal{T}$ and a list of the chords left out of the tree. The corresponding fundamental cycle basis has a cycle for each chord. Orient the cycles in the same direction as their chords, so that the cycle crosses the chord in its forward direction (from the low indexed endpoint to the high indexed endpoint). Then, to construct the curl we need to be able to list the edges from the spanning tree used to complete the cycle associated with each chord. This is the motivation for storing the parent edges, and depending on the chosen implementation, the ancestral edges.

Suppose that we do not store the ancestral edges. Then to find the cycle associated with a particular chord start two searches. The first is initiated at the initial node in the chord, the second is initiated at the final node in chord. Then, using the list of parent edges and trace backwards towards the root of the tree from each node. It is easy to move backwards in the tree given the edge to endpoint mapping M and list of parent edges. For

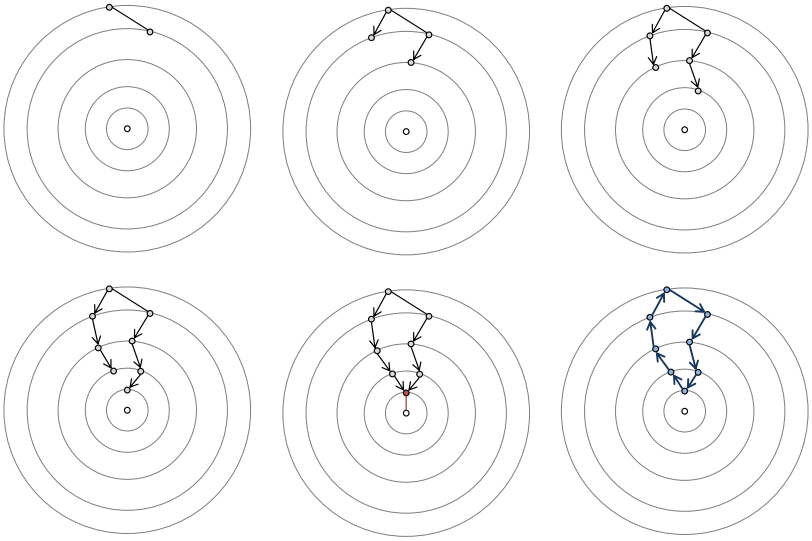

Figure 3.18: Search procedure for finding the loop associated with a chord. The two search paths are shown in black. The intersection and tail are shown in red. The final loop is shown in blue.

node j find the parent edge of node j, $k(j)$, then the parent node is whichever entry of $M(k(j), :)$ doesn't equal j. Since there is only one parent edge for each node in the tree the two searches are each a path. Eventually the two searches must intersect. When they intersect remove any portion of the two searches that overlaps. This produces the cycle associated with the chord as illustrated in Figure 3.18. It is clear that the length of paths in the tree will determine the time it takes to construct the curl. If the tree is tall then the loops are long since the loops only terminate at points where the tree splits. If the loops are long then each search takes longer. If the network is sparse then the number of edges scales in the number of nodes, so the number of loops $L = E - V + 1 \approx (d-1)V$ also scales in the number of edges. As a result, if the average loop has perimeter P then the cost of constructing each loop is $\mathcal{O}(PV)$. If we use a depth first search then $P \approx V$ so the runtime is $\mathcal{O}(V^2)$. Alternatively, if we use a breadth first search $P \approx \mathcal{O}(\log_d(V))$ for

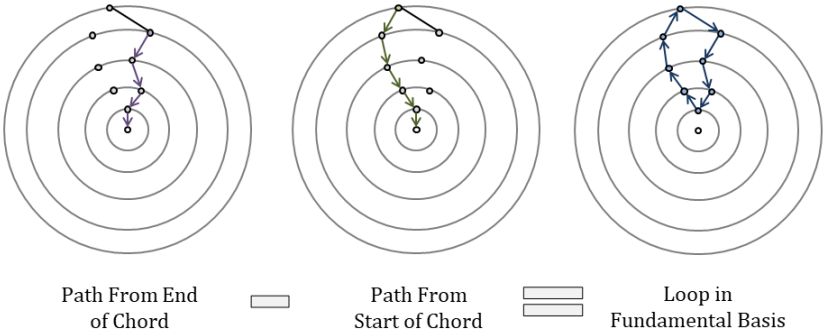

Figure 3.19: The procedure for generating loops by subtracting the list of ancestral edges from the end of the chord from the list of ancestral edges from the beginning of the chord.

sparse random graphs [84] so the runtime is $\mathcal{O}(V \log_d(V))$. The expected maximum loop perimeter depends on the expected distance between randomly chosen nodes, and can be bound above by the expected diameter of the graph. The distribution of graph diameters and the expected distance between randomly chosen nodes are studied for a variety of random graphs in [84, 85, 86, 87, 88]. In particular, if the expected number of degree diverges to infinity slower than $\log(V)$ as $V \to \infty$ then the diameter converges in probability to a finite constant times $\log(V)$ with base equal to the expected degree, and the expected degree is converges to a constant that is strictly greater than 1 then the diameter is $\mathcal{O}(\log_d(V))$ where d is the expected degree [85]. Moreover the fraction of all graphs with fixed V and E such that the diameter is greater than $c \log(V)$ for a fixed constant c converges to zero as V diverges [83, 89].

If storage is not an issue than the list of ancestral edges can be used to construct the curl directly from the list of chords without any additional searching. This direct approach is marginally faster when storage is not an issue, and is easier to implement since it doesn't require a method for detecting the overlap of two paths and removing extra edges.

Consider a particular chord, k. Then $M(k, :)$ stores the endpoints of the chord. Then the difference of the path from the start of the chord to the root and the path from the end of

the chord to the root form the desired cycle (see Figure 3.19). Therefore we can construct the curl by subtracting two matrices from each other, one with all the paths from the start of the chords to the root, and one with all the paths from the ends of the chords to the root. This procedure does not require searching backwards from the endpoints of each chord, but does require extra operations to cancel overlapping paths. Generally these extra operations are not particularly expensive since the length of the paths typically scales in $\log_d(V)$ [84] (or slower [87]).

The paths back to the root from an arbitrary vertex is the list of ancestral edges of that vertex. So, for a chord with endpoints i and j we need only recall the ancestral edges of i and j. We then define two separate lists. One list records all the paths from the starting node of each chord to the root. The other lists all the paths from the ending node of each chord to the root. These lists contain the the edge indices associated with these paths (lists of ancestral edges). To distinguish which sets of indices in each list are associated with each loop we add $(j-1)E$ to the edge indices for the j^{th} loop. Since there are only E edges no edge has index greater than E. Therefore we can distinguish which loop is associated with any element of the two lists by dividing the value by E and rounding down.

Once we have generated these two lists we produce two sparse vectors with LE entries each, and with value 1 at all entries corresponding to a number in the appropriate list. Then the vectors are reshaped into two $L \times E$ matrices. Since we added $(j-1)E$ to all the entries of the lists corresponding to the j^{th} loop, those entries map to the j^{th} row of the reshaped matrices. We then subtract the second matrix from the first (adding in the chords) to produce the curl.

When tested both of these algorithms numerically and found that they typically ran in time $\mathcal{O}(V \log_d(V))$ for sparse random connected networks. A sample of the runtimes and average length of the basis loops for a sequence of random networks from size 10 to 10^4 is

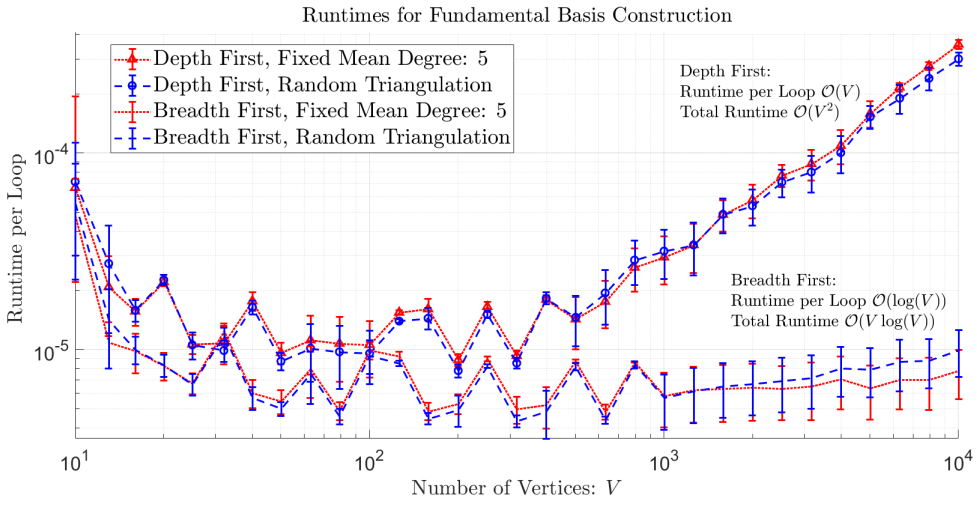

Figure 3.20: Time to construct the curl per loop in the curl using a depth first search and a breadth first search. One hundred random networks were sampled per network size, V. Results for random networks with a fixed average degree of 5 are shown in red, and a Delaunay triangulation of V nodes drawn uniformly from the unit box $[0,1] \times [0,1]$ are shown in blue.

shown in Figure 3.20, and is compared to the equivalent runtimes is a depth first search is used instead of a breadth first search.

Note that the run time per loop using a depth first search is proportional to the number of nodes in the network, while the run time per loop is close to constant. For larger networks the run time using a breadth first search scaled in $\log(V)$. Therefore the overall runtime using depth first search was observed to be $\mathcal{O}(V^2)$, while using breadth first search was $\mathcal{O}(V\log(V))$.

Th difference in run time per loop reflects the average size of the basis loops produced by each method. The average length of the basis loops is shown in Figure 3.21. Note that the average loop produced using a depth first search is longer than the average loop using a breadth first search. Using a depth first search the loop lengths scale in $\mathcal{O}(V)$, while using a breadth first search the loop lengths scale in $\mathcal{O}(\log(V))$. For example, the average length of the loops using a breadth first search with fixed average degree 5 fits to $\log_{5/2}(V) + 0.74$

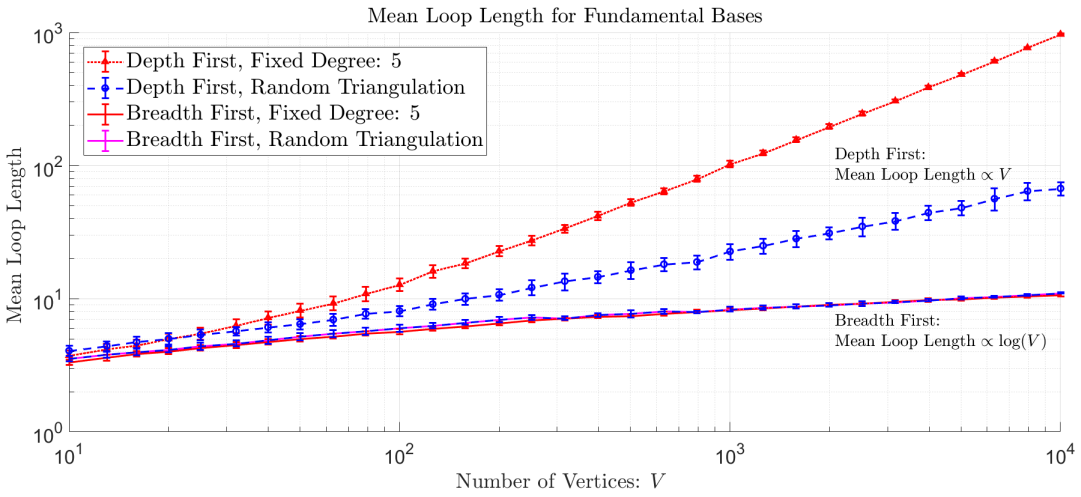

Figure 3.21: Length of the average cycle in the fundamental cycle bases built using a depth first search and a breadth first search. One hundred random networks were sampled per network size, V. Results for random networks with a fixed average degree of 5 are shown in red, and a Delaunay triangulation of V nodes drawn uniformly from the unit box $[0, 1] \times [0, 1]$ are shown in blue.

with R square 0.9994. The length of the loops in a fundamental basis are controlled by the average distance between the endpoints of the chords through the tree. The average distance between two random vertices through the tree generated by a breadth first search is typically smaller than in a depth first search. Since the cost of constructing a loop in the basis is governed by the length of the loop, building a fundamental basis using a breadth first search is cheaper than using a depth first search.

We use a breadth first search to construct fundamental bases from now on since it is faster, produces smaller average loops, and, as a consequence, reuses edges less on average.

3.5.3 A Search Procedure for Constructing a Weakly Fundamental Cycle Basis

In Section 3.5.2 we introduced a simple search procedure for finding a fundamental cycle basis given only an adjacency structure. A breadth first search is used to build a spanning tree for the network. The chords left out of the spanning tree are stored. Then to construct a loop from each chord we search backward through the tree from each endpoint to find the point where the two paths from the endpoints of the chord first overlap. Here we show that, after a simple modification, a similar search procedure can be used to construct a weakly fundamental cycle basis with small loops.

The procedure begins in the same way. The vertices are ordered in decreasing order of degree and the node with maximal degree is set as the root of a tree. Then a breadth first search is performed out from the root. The search is ordered so that the neighborhood of leaves with large degree are always searched before the neighborhood of leaves with small degree. As the tree is constructed all chords are recorded.

Order the chords by the average distance of their endpoints from the root of the tree. This introduces an ordering on the basis cycles. Since a breadth first search is used to build the tree no edge connects any endpoints whose distance from the root differs by more than one. Therefore, if the edges are ordered by the mean distance of their endpoints to the root, then the edges are ordered into groups so that we always consider all of the edges between nodes a distance d from the root before edges between nodes distance d and $d+1$, before edges between nodes a distance $d+2$ from the root, etc.

Consider the loop l corresponding to the l^{th} furthest chord from the root. Let the endpoints of l be $i(l), j(l)$. Then search through a subset of the edges of the full graph for a path from $j(l)$ back to $i(l)$. To construct a fundamental basis this subset was constrained to

the edges of the spanning tree. Suppose that we search backwards through $\mathcal{T} \cup \left(\cup_{j=1}^{l-1} \mathcal{C}_j\right)$. That is, search backwards through both the spanning tree and the chords of any loops that have already been added to the basis. Since the loops are ordered by the average distance of their endpoints from the root, if $i(l)$ and $j(l)$ are distances d and $d+1$ from the root, then we are searching over a subgraph containing the every node and edge within a distance d from the root. The same is true if the endpoints of the chord are both a distance $d+1$ from the root. As soon as the two searches intersect a loop is formed. By expanding the subgraph searched each time a loop is added we increase the possibility of finding a short loop. By construction this loop is one of the shortest loops that can be formed using the specified chord in $\mathcal{T} \cup \left(\cup_{j=1}^{l-1} \mathcal{C}_j\right)$.

The breadth first searches leaving each endpoint may intersect at multiple nodes simultaneously if there are multiple shortest loops involving the chord and $\mathcal{T} \cup \left(\cup_{j=1}^{l-1} \mathcal{C}_j\right)$. Choose one of these loops, denote it $\mathcal{C}_l$, and add it to the set of cycles $\{\mathcal{C}_1, \mathcal{C}_2, ...\mathcal{C}_l\}$. If the number of times each edge is used in the cycle basis is tracked then the cycle could be chosen to minimize the total reuse of edges. This process will necessarily produce a weakly fundamental cycle basis, with one cycle for each chord, since each time a cycle is added to the set it uses at least one edge used by no other cycles. In addition, since the cycle were associated with the chords, the ordering of the chords fixes the order in which the boundary cycles should be removed to perform the iteration in Theorem 13 to recover θ. Moreover, by construction each cycle added is as small as possible using only the edges in $\mathcal{T} \cup \left(\cup_{j=1}^{l-1} \mathcal{C}_j\right)$. This keeps the cycles in the cycle basis small. Finally, each cycle $\mathcal{C}_l$ is chosen to minimize the reuse of edges from $\cup_{j=1}^{l-1} \mathcal{C}_{l-1}$ given that the cycle is as short as possible, so the basis will avoid overusing edges when possible.

By construction every cycle in a weakly fundamental cycle basis constructed using the algorithm described above must have length less than or equal to the cycle formed by the

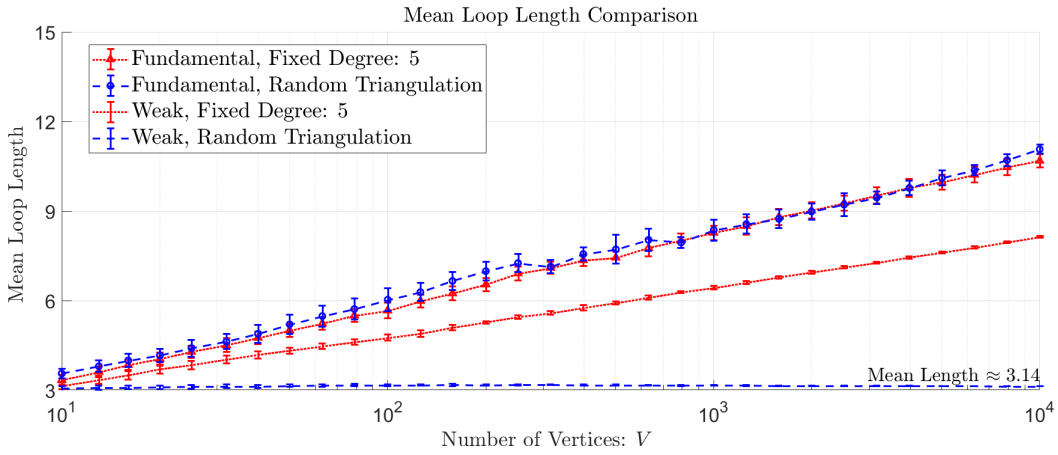

Figure 3.22: Length of the average cycle in fundamental and weakly fundamental cycle bases using a breadth first search. One hundred random networks were sampled per network size, V. Results for random networks with a fixed average degree of 5 are shown in red, and a Delaunay triangulation of V nodes drawn uniformly from the unit box $[0,1] \times [0,1]$ are shown in blue. Note that the length of the cycles produced in the weakly fundamental cycle basis are always shorter than the cycles produced in the fundamental basis, and that for the triangulation the average cycle length is close to 3. The minimal cycle basis for the triangulation (the triangulation itself) would have average cycle length equal to 3.

same chord in a fundamental cycle basis based on the same tree. The average cycle length using a fundamental and weakly fundamental basis are compared in Figure 3.22. If the average degree of the vertices is fixed to 5 then, using a fundamental basis the length of the cycles fits to $\log_{2.5}(V) + 0.74$ (r-squared equal to 0.9994). Using a weakly fundamental basis the length fits to $\log_4(V) + 1.465$ (r-squared equal to 0.9998). Thus, for large enough V, the cycles produced using a weakly fundamental basis are about 1.5 times shorter than the cycles formed using a fundamental basis.

The difference between the two bases when applied to the random triangulations is more striking. For a random triangulation the minimum cycle basis should have average cycle length equal to three since the set of all triangles is a planar basis for the graph. Moreover, since the graph is planar no edge should be used by more than two cycles, so the

average number of times any edge is reused should be less than two.

The length of the cycles in the fundamental basis increase at the same rate when applied to the random triangulations as when applied to the random graphs with average degree fixed. Thus, when using a fundamental basis the cycle lengths grow logarithmically in the number of vertices.

In contrast, when the weakly fundamental basis is used the average cycle length remains close to 3, peaking at $V \approx 300$ with average length 3.18, then decays back towards 3, reaching an average of 3.13 at $V = 10^4$. It follows that the weakly fundamental cycle basis is close to minimal. The average number of times any edge appears in the basis reaches a maximum of 2.1 ± 0.01 for $V \approx 600$, and the average over the ensemble of graphs of the maximum number of reuses approach 5.5 ± 0.5, indicating that the weakly fundamental basis is also close to a planar basis. In contrast the average edge reuse and maximum edge reuse increased monotonically in V if a fundamental basis was used, with the maximum crossing 100 reuses at V on the order 10^3.

The weakly fundamental search algorithm is very efficient for graphs with many neighboring small loops, since each search only continues until a loop is found. In fact, when tested on a sequence of random Delaunay triangulations with 10 to 10^5 vertices this method ran in $\mathcal{O}(V)$ time, with an average cycle length of 3.14 for large networks.

The run time of this search is approximately linear in V if the length of loops found and average degree of each node is independent of the total number of nodes in the network. Then the depth of each loop search is independent of the number of nodes, and the number of nodes searched within that depth is approximately constant. If the largest cycle found by this method has perimeter P and the average degree of each node is d then the expected number of nodes searched by each loop search is roughly $2d^{\lceil P/2 \rceil}$ since the largest tree leaving the endpoint of each chord will have diameter less than or equal to $\lceil P/2 \rceil$. In

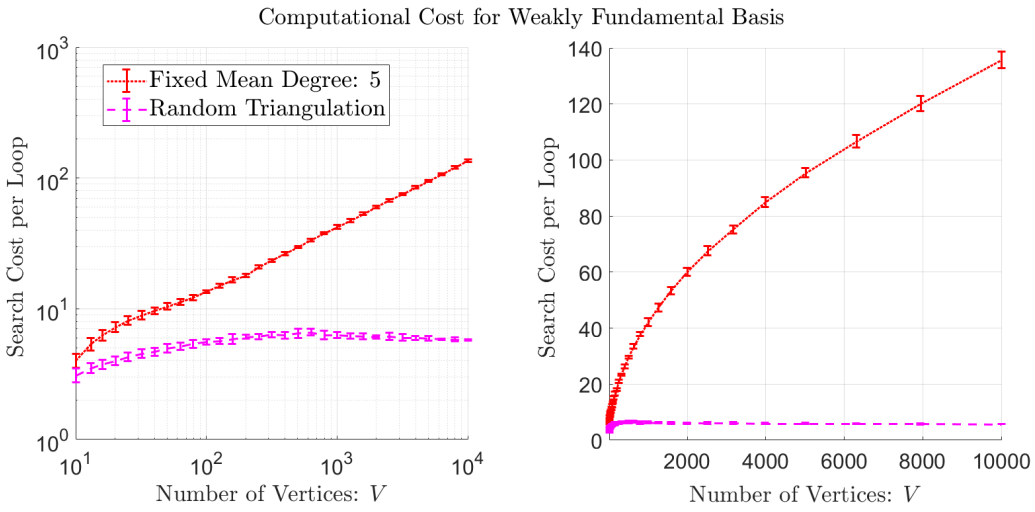

Figure 3.23: The computational cost per loop to construct a weakly fundamental cycle basis. One hundred random networks were sampled per network size, V. Results for random networks with a fixed average degree of 5 are shown in red, and a Delaunay triangulation of V nodes drawn uniformly from the unit box $[0,1]\times[0,1]$ are shown in blue. The left panel is a log-log scale, and the right panel is linear. Note that the computational cost per loop is approximately constant for the random triangulation, but increases $\mathcal{O}(V^{1/2})$ for random networks with a fixed average degree (increases with slope $1/2$ on the log-log plot for $V > 10^2$).

Figure 3.23 the computational cost per loop of building a weakly fundamental cycle basis on a random triangulation is approximately constant in V since the cycle lengths remained near 3, and beneath 6, for $V \in [10^2, 10^4]$. The exact cost of a weakly fundamental basis search for a random graph which finds cycles with average cycle length P depends on the degree distribution and clustering of the network since it depends on the average size of the search trees used to complete each basis loop.

In contrast, if the average loop size grows with V then the computational cost per loop will increase as nodes are added. In Figure 3.23 the computational cost per loop of building a weakly fundamental basis for uniformly sampled random graphs with fixed average degree increased proportional to $V^{1/2}$ (fits to $1.1V^{0.51} + 1.32$ with r-squared 0.9991). This scaling can be argued directly from the expected sizes of the spanning trees used to find the

loop associated with each chord. The average path length between any pair of randomly chosen endpoints increases proportional to $\log_d(V)$ where d is the average degree. Then, since the edges are drawn uniformly, the expected length of the shortest path segment connecting two endpoints of a chord is approximately $\log_d(V)$. Since the loops are found by a pair of searches starting from each edge the length of the search trees is less than $\lceil \log_d(V)/2 \rceil$. Then the size of each search tree is proportional to $d^{\lceil \log_d(V)/2 \rceil} = \mathcal{O}(V^{1/2})$. If the average degree of each node is fixed then the number of loops grows proportional to V, so the computational cost for building the weakly fundamental basis is $\mathcal{O}(V^{3/2})$.

Thus, for networks with small loops, or with a high clustering coefficient, the cost of building a weakly fundamental cycle basis may be close to constant per loop, while for networks without any tendency to cluster, the computational cost per loop may scale in the square root of the number of nodes.

Kavitha provides bounds for the minimal total length of fundamental and weakly fundamental bases. Every graph has a weakly fundamental basis of length $\mathcal{O}(E \log(V)/\log(\frac{E}{V}))$ thus there is a minimal weakly fundamental cycle basis with total length that scales slower than $E \log(V)$ in the graph size. In contrast, the equivalent bound for fundamental bases is $\mathcal{O}(V^2)$. Finding a minimal fundamental or weakly fundamental cycle basis is APX hard[6] [33] Depending on the application it may be enough to work with the weakly fundamental basis produced by the search procedure outlined in this section. If the resulting basis is not sufficiently minimal then the length of the basis can be reduced by searching for pairs of overlapping cycles in the basis who could be combined to form a pair of shorter cycles.

[6] A problem is APX hard if there is no polynomial time approximation algorithm unless $P = NP$.

3.5.4 Special Case: Planar Graphs

Suppose $\mathcal{G}$ is planar and biconnected[7] How can we find a planar cycle basis for $\mathcal{G}$ using only its adjacency structure?

Let $\mathcal{F}$ denote the faces of the graph after some embedding. Then $\mathcal{F} \setminus \mathcal{F}_j$ is a planar basis for any $j \in [1, L+1]$. Whichever face is not included is the exterior face of the graph in some embedding. It follows that, if we can identify the faces of the planar graph, then we can easily construct a cycle basis by choosing a face to consider as the exterior, and setting the cycle basis equal to all other faces of the planar graph. For planar graphs a minimum planar basis consisting of the interior faces can be found in linear time by solving an all-pairs min-cut problem on the dual, which runs in $\mathcal{O}(V^2 \log(V))$, or $\mathcal{O}(V^2)$ time depending on the implementation [90, 36, 33].

Suppose we find a cycle $\mathcal{C}$ in $\mathcal{G}$. How can we tell if it is a face? If the cycle, $\mathcal{C}$, is a face of the planar graph then the graph formed by removing all edges and nodes from the cycle, $\mathcal{G} \setminus \mathcal{C}$, is still connected. If, on the other hand, the cycle is not a face, then after removing the cycle the graph will be broken into two or more disconnected components - and the separated components can be sorted into the interior and exterior of the cycle [36]. It follows that a cycle is a face if an only if ot can be removed from a planar graph without separating the graph into two components.

This classification rule introduces an intuitive algorithm for finding all of the faces of a planar graph. The algorithm is based on recursively partitioning the planar graph into smaller graphs. In contrast, all of our algorithms thus far have been edge insertion algorithms, in which edges are added one at a time to build the cycle basis. It is possible to build a planar cycle basis and embedding by adding edges or nodes one at a time

[7] If $\mathcal{G}$ is singly connected then isolate all singly connected components and apply the method to each singly connected component separately

(cf. [91, 92]), however these algorithms require careful ordering of the edges [32]. The partitioning approach is inspired by Hopcroft's planarity test [32] and Rusoz' algorithm for enumerating all of the cycles of a planar graph [37]. First we present a graphical algorithm in which the partitioning is performed explicitly. Then we present a modified version of the partitioning algorithm that uses a depth first search to avoid searching over the same subgraphs repeatedly.

Initialize a depth first search from some node in the graph. Continue the depth first search until we reach a node who only neighbors nodes we have already added to the tree. This is the first leaf in the tree. As the search progresses store all of the chords. If the search is stopped at the first leaf then the tree is a path, and any cycle formed by a chord consists of the chord, plus all edges in the segment of the path between the endpoints. Thus the length of every cycle can be computed by indexing the nodes in the order they are discovered by the search and subtracting the lower endpoint index from the larger endpoint index of each chord. Everytime a chord is added compute the length of the associated cycle and compare it to the length of the largest cycle found thus far. If it is the larger then update the length of the largest cycle and store that chord as the chord corresponding to the longest cycle. Thus, once the first leaf of the tree is found we can easily identify the largest cycle formed by a back-edge along the first branch of the spanning tree.

Form a new graph $\mathcal{G} \setminus \mathcal{C}$ by removing all the nodes and edges in the cycle from the graph. Then check whether the resulting graph is connected. If it is connected then the cycle is a face. If it is not connected then the cycle is not a face. To check if the graph is connected start a search from a node in $\mathcal{G} \setminus \mathcal{C}$. If the search reaches every node in $\mathcal{G} \setminus \mathcal{C}$ then the graph is connected. The cost of the search is strictly less than $V + E$ since every vertex and most of the edges in $\mathcal{G} \setminus \mathcal{C}$ will be checked once.

Suppose that the cycle is a face. Then, by convention, we will let it be the exterior

face of the graph. Suppose the cycle is not a face. Then $\mathcal{G} \setminus \mathcal{C}$ can be separated into two components, an interior and an exterior. This sorting can be accomplished using the method presented in [32]. We skip the implementation details since this version of the partitioning algorithm is included to provide graphical intuition for the improved algorithm discussed at the end of the section.

Let $\mathcal{G}_1$ denote the union of the first component with the cycle $\mathcal{C}$, and let $\mathcal{G}_2$ denote the union of the second component with $\mathcal{C}$. Then both $\mathcal{G}_1$ and $\mathcal{G}_2$ are planar graphs, and $\mathcal{C}$ is a face of both graphs. By convention let $\mathcal{C}$ be the exterior face for both of the new graphs. The motivation for starting by searching for a large loop is to attempt to subdivide $\mathcal{G}$ into two graphs, each approximately one-half the size of $\mathcal{G}$. In either case we are left with a set of planar graphs with a cycle that is the exterior face of each planar graph.

A bridge is a path through a planar graph that intersects two different nodes in the exterior face without using any of the edges of the exterior face. Pick an initial node in the exterior face and search out from the node without using any of the edges in the exterior face. Once another node from the exterior face is reached the path back from that node to the starting node in the search tree is a bridge. The union of the bridge with the exterior face forms a pair of neighboring cycles that share the bridge. Denote these cycles $\mathcal{C}_1$ and $\mathcal{C}_2$. As before, identify whether or not each cycle is a face, and if it is a face store it in the list of faces found. If the cycle is not a face then the planar graph can be subdivided into two smaller planar graphs exactly as before, and the exterior face of the two graphs are $\mathcal{C}_1$ and $\mathcal{C}_2$. Then the process can be iterated. At each stage the goal is to pick a bridge that separates the planar graph into two components of approximately equal size.

Thus the procedure is as follows. Identify a large cycle $\mathcal{C}$. Remove the cycle from the original graph and identify any connected components. If the remaining graph is connected ($\mathcal{C}$ is a face) then let $\mathcal{G}_1$ equal $(\mathcal{G} \setminus \mathcal{C}_1) \cup \mathcal{C}_1 = \mathcal{G}$ and let $\mathcal{G}_2 = \emptyset \cup \mathcal{C} = \mathcal{C}$. Then both $\mathcal{G}_1$ and

$\mathcal{G}_2$ are planar and the exterior face of each is $\mathcal{C}$. Then, for each planar graph that is not a face find a bridge connecting two different nodes in the exterior face. Form a cycle from the bridge and the exterior face. If no such bridge can be found then find a loop starting from a node in the exterior face. Then iterate the procedure used for the first cycle. Anytime a face is identified add it to the list of faces and keep a tally on each edge of how many of the identified faces it is included in. Anytime an edge is used in two faces it can be removed from the graph. An example is illustrated in Figure 3.24

This process iteratively partitions the planar graph into smaller and smaller components until all of the faces are identified. Once all the faces are identified a planar basis can be constructed by removing one face from the list.

The computational cost of this partitioning algorithm depends on how close each bridge comes to halving each subgraph. At each step a bridge is used to split a graph into two parts. Therefore the algorithm produces a binary tree of graphs with one leaf for each face in the graph. There are $L+1$ faces so the tree so has $L+1 = E-V+2$ leaves. A binary tree with n leaves always contains $2n-1$ vertices, so the algorithm always runs in $2(L+1)-1 = 2L-1$ steps. The cost for each step depends on the size of the graph treated in the step. The cost to search for a bridge is strictly less than the cost to search the entire subgraph, which is linear in the number of vertices plus the number of edges. Once the bridge is found the cost of identifying exterior and interior components is also linear in the size of the subgraph [32]. A connected component can be found by a search procedure. Thus the computational cost of each partitioning step is linear in the size of the subgraph partitioned.

Now consider the total cost of partitioning every graph in the list of subgraphs generated at a particular stage. If no faces have been identified yet then the union of the set of components is the original graph. The components may overlap at the nodes and edges

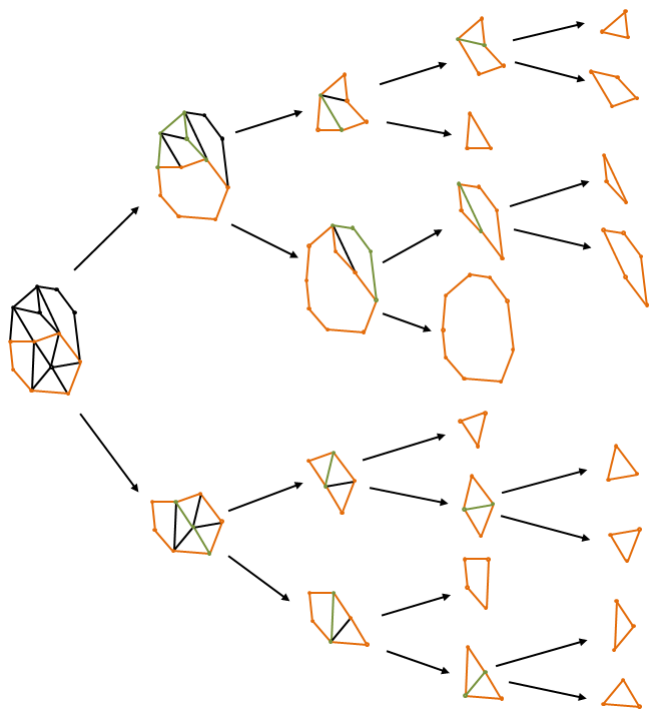

Figure 3.24: Constructing a planar basis by recursively partitioning the graph. In the first stage a loop is found. The loop is shown in orange. The graph is then split into two planar graphs. The loop is a face of each of the two components. Then a bridge is found connecting two nodes in the face. The bridges are shown in green. The bridge is used to separate the graph again. The loop formed by the bridge and a half of the original exterior face is an exterior face for the two components. This process is repeated recursively until all 12 faces of the original graph, including the exterior face, have been found. Note that the tree has 12 leaves, one for each face, and 11 junctions where a planar graph is partitioned into two smaller planar graphs. Also note that the height of the tree is 4, which is the minimum possible height since $2^4 = 16$ is the smallest factor of 2 greater than or equal to 12.

in the exterior cycles. By construction no node or edge is ever included in more than two components, so the sum of all the nodes and all the edges in all of the subgraphs is strictly less than twice the sum of all the nodes and edges in the original graph. Once a face has been found it becomes a leaf in the binary tree of components, so if faces have already been found then they do not contribute to the sum. Thus the cost of each stage is linear in the size of the original graph, and the overall cost is less than the product of the height of the

binary tree with the size of the original graph.

The number of stages depends on the height of the binary tree. If, at every stage, one of the exterior cycles identified is a face then the binary tree includes $L + 1$ stages, so has height $L + 1$. Then the overall cost will scale quadratically in the size of the original graph. In contrast, if the number of faces is a power of two, and every exterior cycle/bridge found splits each component so that half of the faces of the component fall into one subcomponent, and half in the other, then the tree has height $\log_2(L + 1)$, so the overall cost will be $\mathcal{O}((E + V) \log_2(L)) = \mathcal{O}((E + V) \log_2(E - V))$. No binary tree with $L + 1$ leaves is shorter than $\lceil \log_2(L + 1) \rceil$, so the minimal computational cost of the partitioning algorithm is $\mathcal{O}((E + V)\lceil \log_2(L)) \rceil$.[8]

The principal cost of performing this algorithm is the repeated need to search for the connected components of a graph after some nodes and edges have been removed. This requires repeatedly searching subgraphs of the same the graph, so is inefficient. An improved algorithm is presented below which avoids researching the same subgraph multiple times. It is inspired by [32], and uses one depth first search to contruct a sequence of cycles that are used to partition the graph.

Start by running the first step of the previous algorithm to completion. Then, either we find a face of the original graph, or we split it into two pieces, and the cycle used to split the graph is an exterior cycle for each component. To find the connected components we

[8] These considerations highlight the importance of finding a bridge which comes as close to halving each component as possible. To guide the bridge search pick an initial node on each exterior cycle, and compute the cumulative sum of the degree of the nodes in the exterior cycle moving around the cycle clockwise, excluding edges between nodes in the cycle. Then run a depth first search on the component without the edges of the cycle, starting from the initial node. If a bridge, or set of bridges was found then pick the bridge which arrives at the node on the exterior cycle which comes as close to halving the sum of the degree, excluding cycle edges, of all nodes in the exterior cycle as possible. Then, since we assumed the graph was biconnected, all edges leaving a node in the cycle that is not an edge in the bridge must be part of a face in one of the components after partitioning the graph. Thus, by picking the bridge to halve the net degree of nodes on either side of the bridge, only counting edges not in the cycle, we hope to halve the graph so that close to an equal number of faces are contained on either side of the bridge.

had to run a search over the original graph with the cycle removed. Implement this search as a depth first search over the components, starting from a node in the cycle, and searching edges in the cycle before any other edges in the component. Then, after the first partitioning stage, we have a spanning tree for each connected component of $\mathcal{G} \setminus \mathcal{C}$. Focus on a particular component. If we can construct an algorithm for finding the faces of a component, then we can run the algorithm on the original set of components separately to find all faces of the original graph.

Therefore, without loss of generality, assume that we start with a planar graph, and with a spanning tree that starts by tracing around the exterior face of the graph. Let $\mathcal{T}$ denote the spanning tree formed by the search. Run the search to exhaustion so that all chords are found. Orient all the edges in the direction they were searched. Then, if we cross any chord backwards, and trace backwards through $\mathcal{T}$ we will necessarily form a cycle. Keep a list for every node of the chords arriving at that node, and keep a list of which chords have been used. Keep a list of cycles, and keep a list of faces found.

As when building a weakly fundamental cycle basis our goal is to use each chord once, adding to the list of cycles each time the chord is added. Mark the first chord as used. Then, using the list of chords arriving at each node, find all chords arriving at nodes in the cycle that have not been used yet. If there are no such chords add the cycle to the list of faces, remove the cycle from the list of cycles. Then pick a new chord and start again. If there is such a chord cross one of them, then trace backwards through $\mathcal{T}$ until we reach the cycle. An example is shown in Figure 3.25.

Now there are two possibilities, either we arrive back at the node we left, or we arrive at a different node in the cycle. If arrive back where we started then we have found a new cycle. Add the new cycle to the list of cycles. If we did not arrive where we started then the path traced back across the chord through $\mathcal{T}$ is a bridge between two different nodes in

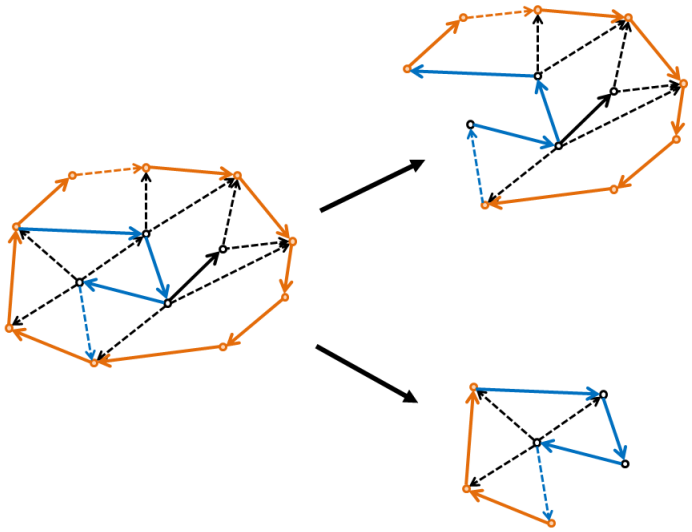

Figure 3.25: Partitioning a cycle into two cycles using a bridge. The solid edges represent the edges in the spanning tree, and the dashed edges are the chords. The arrows point in the direction of the search. The orange cycle is partitioned into two cycles by the blue bridge. The bridge is found by crossing a chord backwards from a node in the cycle, then tracing back through the spanning tree to the cycle.

the cycle. Then the cycle can be split into two cycles, both sharing the bridge. Remove the original cycle from the list and add both of these new cycles to the list of cycles.

This initializes a list of cycles. Pick the first cycle in the list. Search for any chords arriving at the cycle that have not been used yet. If there are none the cycle is a face. Add it to the list of faces, and remove it from the list of cycles. If there is a set of chords arriving at a node in the cycle which has not been used yet then pick one and cross it backwards. Mark the chord as used. As before, trace backwards through the tree until we arrive back on the cycle. If we arrive back at the same node we left we have found a new cycle. Add the current cycle and the new cycle to the end of the list of cycles, and remove the current cycle from the front of the list. Alternatively, if we found a bridge, form two cycles using the bridge and the current cycle, remove the current cycle from the front of the cycle list, and add the two new cycles to the back. If, at some point the list of cycles is empty but not

all chords have been used then pick an unused chord and start the process again. Continue until all chords have been used and the cycle list is empty. If the all chords are used and there are still cycles in the cycle list remove them from the cycle list and add them to the list of faces.

How many faces does this method produce? Every face in the list of faces starts as a cycle in the list of cycles. Every cycle in the list of cycles is added after crossing a chord. When a chord is crossed we either find an entirely new cycle, or partition an existing cycle into two cycles. In the first case one cycle is added to the list, in the second two are added and one is removed. Thus, every time a chord is crossed the cycle list gains one new cycle. Cycles leave the cycle list once there are no chords arriving on the cycle that have not been used. When a cycle is removed it is added to the list of faces, so, moving a cycle from the cycle list to the face list does not change the sum of the lengths of the two lists. Thus, the sum of the length of the two lists increases by exactly one cycle every time a chord is used. Since there are L chords, and both lists start empty, after all L chords are used the sum of the length of the two lists is L. Then all of the cycles in the cycle list are moved into the face list, so the method will always produce L faces.

The set of L faces is a set of L cycles. To show that they form a planar cycle basis it is enough to show that they form a cycle basis, and that no edge is ever used in any cycle more than twice. The latter is easy to show directly from the structure of the algorithm. Edges only enter the set of cycles when a chord is crossed. If they are part of a new cycle then they appear once. If they enter as part of a bridge they appear twice. Once an edge enters it remains part of at most two cycles, since, if an edge is part of a cycle then moving that cycle to the set of faces does not increase the number of times the edge is used, adding an entirely new cycle to the set of cycles adds an entirely new list of edges so does not increase the number of times the cycle is used, and partitioning a cycle by adding a bridge splits the

cycle into two parts before adding new edges, so does not change the total number of times any edge already in the cycle is used. Therefore no edge appears in the set of L faces more than once.

All that remains is to show that that the set of L faces are a cycle basis. Consider the set of cycles formed by appending the list of faces to the list of cycles at a intermediate stage in the algorithm. Then a curl operator could be constructed associated with that set, and the number of independent cycles in the set would be the rank of the curl. Now, when a chord is crossed either one row is removed and two are added (a bridge is found), or a new row is added to the end (an entirely new cycle). If a new row is added it includes an edge no other row includes, the chord, so must be independent of the previous rows. If a bridge was found then an existing cycle is split in two, the original cycle is removed, and the two new cycles are added. This can be represented by first finding one cycle that uses the bridge and adding a row associated with it. This row includes an edge no other row uses, the chord, so is independent of all previous rows. Then subtract that row from the row associated with the original cycle, and move it to the end. These are elementary row operations so do not change the rank of the curl. Thus, whenever a chord is crossed the rank of the curl increases by one. Moving a cycle from the list of cycles to the list of faces only permutes the rows so does not change the rank. After L chords have been crossed the curl has rank L, so the set of cycles form a cycle basis.

Therefore this cycle partitioning algorithm is guaranteed to produce a planar cycle basis. The algorithm is more efficient than the graph partitioning algorithm since it only requires one initial search which has cost order $\mathcal{O}(E + V)$, then the algorithm has one stage for every chord, so always runs in L stages. For each stage we trace back through a spanning tree, then through a cycle to form a loop. This has an expected cost equal to the length of the cycle formed by the back trace. If this cycle only uses edge from the tree then it is

added to the end of the list of cycles. If, instead, the cycle is formed by combining part of the original cycle with a bridge then two new cycles need to be formed, both sharing the bridge, so the cost is twice the length of the bridge plus the length of the original cycle. All other operations simply involve permuting rows of the curl, so the overall computational cost will depend on the average cycle size formed each time a chord is added, and how quickly these cycles shrink as the graph is partitioned

The cycle partitioning algorithm is an example of a backtracking algorithm [37]. Backtracking algorithms are widely used to enumerate the cycles of graphs. For example, [93] presents a time optimal algorithm for enumerating all cycles of a graph via backtracking [37]. Hopcroft's planarity test [32] is an example of a backtracking algorithm. Backtracking algorithms usually cannot be efficiently parallelized. An advantage of working with a binary partitioning is that, at each stage, the component graphs/cycles can each be partitioned in parallel. Rusoz presents a parallelizeable algorithm for finding all cycles of a planar graph in $\mathcal{O}(V^2)$ time [37]. Another alternative to the approach developed here is to use a planar graph drawing algorithm to identify the faces.

3.6 Summary

In Section 3.2 we illustrated the decomposition for a sequence of simple networks. These included trees, loops, networks with isolated loops, an example with a pair of linked loops, and complete networks. For loops, or networks with isolated loops, the rotational potential and flow can be recovered by evaluating the curl around each loop, then dividing by its perimeter. For complete networks the decomposition can be performed directly by computing the divergence of f at every node, then dividing by the number of nodes in the network. This shortcut will be used in Chapter 4 to compare the HHD to existing ranking methods defined on complete graphs.

In Section 3.3 we developed a framework for performing the HHD on a graph formed by the products of smaller graphs. This framework offered insight into the cycle space of lattices and hypercubes, and provided explicit construction rules for building the operators of product graphs using Kronecker products of the operators defined on the factor graphs (see Section 3.3.2). The framework also led to an elegant spectral method (Section 3.3.3) for performing the HHD on lattices that reduces the cost from $\mathcal{O}(V^2)$ to $\mathcal{O}(V \log(V))$ by using an FFT. The implementation details are described in Section 3.4.3.

In Section 3.5 we developed general purpose numerical methods for performing the HHD given only the adjacency structure of the network and the edge flow. We showed that construction of the gradient and node Laplacian is trivial, and that the potential, conservative, and rotational flows can be computed easily using classical techniques once the gradient is formed. The rotational potential depends on the chosen cycle basis. The cycle basis is not unique, and depending on the application some cycle bases may be more appropriate than others. Three search algorithms were presented that are designed to find a fundamental cycle basis, weakly fundamental cycle basis, and planar cycle basis.

Part III

Structure: Application to Tournaments

Chapter 4

Application to Tournaments: Theory

4.1 Preface

This chapter is adapted from a paper submitted to SIAM Review on cyclic competition in tournaments. Some of the derivations presented in Chapter 2 are repeated here since this chapter is designed to be self-enclosed. The derivations are included to show how the HHD can be motivated in an applied setting.

This chapter applies the HHD to competitive tournaments. The HHD is used to simultaneously rank and rate competitors, to identify competitive cycles, and to quantify how cyclic competition is. We compare the HHD to existing methods. The comparison shows that the HHD is conceptually similar to existing methods, but computationally simpler, and offers a more unified analytic approach. The chapter concludes by introducing trait-performance models. A trait-performance model is a statistical model for sampling edge flows. The edge flow is assumed to model the perfromance of competitors, and is a function of the competitors' traits, which are sampled from a trait distribution. We study

the expected sizes of transitive (conservative) and cyclic (rotational) competition when the edge flow is sampled from a trait-performance model. We show that the expected size of the cyclic and transitive components are controlled by the density of the network, and by the correlation in the edge flow on pairs of edges that share an endpoint.

4.2 Introduction: Tournaments, Ranking, and Intransitivity

Competitive tournaments are important across disciplines. Examples range from ecology and animal behavior [94, 95], to psychology and sports [96, 97]. Rating and ranking is important in each of these areas. In sports, ranking and rating teams and players is a topic of broad popular interest. Rating is important in biology since fitness is an intrinsic rating of competitive ability since survival and reproduction are influenced by repeated competitive interactions with many individuals. Ranking is especially important in politics, as many electoral systems determine a winner by aggregating votes into a partial ranking of the candidates. Ratings and rankings are often sought since they simplify the description of a tournament by assigning each competitor a single number that purports to measure how good they are.

Not all tournaments allow for a consistent ranking of competitors. This observation motivates classification into transitive and intransitive tournaments. A tournament is *transitive* if knowing that A usually beats B, and B usually beats C, is enough to conclude that A usually beats C. Transitive tournaments are consistent with a global ranking of all the competitors. An *intransitive* tournament is a tournament that is not consistent with any global ranking. Intransitive tournaments must contain at least one cycle where the

transitive assumption fails. Examples of intransitive tournaments appear in practically every discipline where tournaments are studied [18, 98, 99, 100, 101], and are the norm rather than the exception when using real data [16, 97, 102, 94, 95, 103, 104]. Intransitivity may arise due to uncertainty in observed data [97, 104], or may be intrinsic to competition as in the game of rock-paper-scissors.

Intransitivity is important for two reasons.

First, intransitivity presents a challenge when ranking competitors since no ranking is consistent with the tournament. For example, Condorcet's paradox is a voting paradox in which voter's preferences lead to cyclic community preferences [98].[1] Because of the cyclic community preferences there is no way to fairly rank the candidates, and, as a consequence, pick a winner of the election.

Second, when intransitivity is intrinsic to the structure of the tournament then the tournament contains cyclic structure, as in rock-paper-scissors. Cyclic structures can radically alter optimal strategies [18] and long term dynamics [99, 101, 105, 106, 107]. For example, in ecology it is widely hypothesized that intransitive competition between species promotes biodiversity since no species dominates. This hypothesis is based on extensive theoretical work [94, 99, 101, 105, 106, 107, 108] and limited case-studies of small species assemblages [109, 102, 110, 111, 103]. However, the importance of intransitivity in real natural communities is controversial [112, 113, 114] - in part because there are few robust metrics for measuring intransitivity from incomplete and noisy data. It has been shown that uncertainty in data can easily be conflated with observed intransitivity, and that common sampling methods for filling in missing data overestimate intransitivity [95].

Thus there is a need for ranking and rating methods that are robust to intransitivity and

[1] Suppose there are three candidates in an election and three voters. Suppose that the first voter prefers A to B to C, the second B to C to A, and the third C to A to B. Then A would beat B in an election between the pair, B would beat C, and C would beat A.

measures of intransitivity that can handle noisy and incomplete data.

Jiang et al introduced the discrete Helmholtz-Hodge Decomposition (HHD) as a general method for ranking objects from incomplete and imbalanced data [16]. The decomposition is a network theoretic tool that we adapt to the study of competitive tournaments. The HHD accomplishes three fundamental tasks. First, it assigns a rating to each competitor. Competitors can be ranked accordingly. Second, it produces a measure of intransitivity that quantifies how far an observed network is from the nearest perfectly transitive network. Third, it represents the observed network as the direct sum of perfectly transitive and a perfectly cyclic networks. This decomposition provides an elegant characterization of intransitivities present in data, and can reveal underlying cyclic tendencies in tournaments. This last property was leveraged by Candogan to identify cyclic structures within collections of competing strategies [18].

When compared to existing ranking methods and intransitivity measures, the discrete HHD is attractive has a number of advantages. It is more general than some classical methods since it applies to arbitrary network topologies and can accommodate imbalanced data [16]. It is also more informative because it provides a clear description of both underlying transitive and cyclic structures. Most ranking methods and intransitivity measures focus on the transitive component while the HHD puts the transitive and cyclic components on equal footing. Finally, it remains efficiently computable even for large, incomplete networks [16]. In contrast, Slater's index [104] requires solving an NP hard optimization problem [115, 116], and Kendall's index [97] requires a complete network.

This chapter aims to answer two fundamental questions:

1. Why use the HHD when other methods exist?

2. Having chosen to use the HHD, what do we expect when pairwise competitive ad-

vantage derives from traits drawn from an underlying distribution?

Answering the first question is important since there are many possible methods to choose from, so the choice of method should be made in a principled way. Answering the second question is important since it builds a conceptual bridge from the competitors and competitive event to the overall structure of tournament. As in Landau [117], we seek to understand how the underlying distribution of traits among competitors, and the relationship between traits and success influence the overall tournament.

This is an important question across disciplines. In biology the relationship between certain traits and success in competition for survival and reproduction is intrinsically related to fitness, and selection for heritable traits [118]. For example, competition for social dominance among male elephant seals depends on their body mass [119] and competition among male dwarf Cape chameleons depends on coloration, head size, and body length [118]. Success in these competition events is correlated with reproductive success, suggesting that heritable traits which improve a male's chances of success are strongly selected for [119]. In sports the relationship between the traits of a player or team and their success is an area of active interest - for athletes, owners, fans, and researchers alike. The rise of sabermetrics, the statistical study of baseball, is a popular example [120]. Sabermetrics have been used to predict the performance of players and teams based on their previous statistics. This includes the prediction of wins and losses as in [121] where it was found that the success of a team depended on a variety of traits including batting average, fielding percentage, slugging percentage, and starting pitcher earned run average.

This chapter answers questions 1 and 2 as follows:

1. Rather than imposing the HHD framework ad hoc, we show that it arises naturally from the study of ranking and intransitivity. To illustrate this point, we provide a

different derivation of the HHD than is provided by [16]. Instead of starting from the decomposition, we propose two special classes of tournaments with clear statistical motivation. We then show that any tournament can be uniquely decomposed into a combination of tournaments from these classes. This decomposition is the HHD (see Theorem 19). Next we illustrate that the HHD can be reached by six different approaches (corollary 19.1), and is thus robust to varying motivations.

2. We show that, under simple assumptions on the distribution of traits, the expected sizes of the components of the decomposition can be computed explicitly from the number of competitors, number of pairs who could compete, and the correlation in the performance of A against B with A against C. This correlation is shown to equal the uncertainty in the expected performance of a competitor. This relation links a decomposition of uncertainty in performance, to correlations in performance, and to tournament structure (see Theorem 20 and corollary 20.1).

The answers to the second question prove, under minimal assumptions, a series of intuitive statements about transitive/cyclic competition that appear, as heuristics, across the literature. These include:

1. (a) The more predictable the performance of A against a randomly drawn competitor (i.e., the less the performance of A depends on their opponent) the more transitive the tournament.

   (b) The less predictable the performance of A against a randomly drawn competitor (i.e., the more the performance of A depends on their opponent) the more cyclic the tournament.

2. (a) The more correlated the performance of A against B with the performance of A against C, the more transitive the tournament.

(b) The less correlated the performance of A against B with the performance of A against C, the more cyclic the tournament.

3. The more pairs of competitors who could compete, the more cyclic the tournament is, on average.

4. Filling in missing data by random sampling overestimates intransitivity.

The chapter is structured as follows. In Section 4.3 we provide some necessary background. Next, in Section 4.5, we derive the HHD in the context of tournaments and develop the associated ratings and intransitivity measure. In Section 4.6 we show how assumptions about the statistics underlying competition promote or suppress intransitivity. We focus on trait-performance models in which performance is assumed to be a function of traits, which are sampled from a trait distribution. We present a theorem (20) which allows the expected size of the intransitivity measure to be computed directly from the number of competitors, edges in the network, and correlation in the performance of A against B with A against C. This result is extended by a corollary (20.1) which shows that the correlation in performance is related to a decomposition in the uncertainty of the performance of A against B. These results lead to a deeper conceptual understanding of how cyclic structure can arise from uncertainty in performance, and can be suppressed by correlation in performance. We present an example to illustrate the explanatory power of this theorem in Section 4.7.

4.3 Background

Consider an ensemble of m competitors. Assume that each competition event involves exactly two competitors, and never results in a tie. This standard assumption [97, 94] can

be weakened to allow for ties. We will refer to competition of this kind as a tournament.[2]

A tournament is specified by a schedule, and a set of win probabilities. The schedule fixes the order of events, and could be either fixed or random. For each possible pairing there is a pair of win probabilities. Let p_{AB} denote the probability competitor A beats B. The shorthand $A > B$ denotes the case when A is expected to beat B ($p_{AB} > 1/2$). It is the direction of competition. In principle the win probabilities could change in time, and could depend on the history of the process. We will focus on tournaments with unchanging win probabilities since evolving probabilities require additional modeling of temporal dynamics (see [122]). In addition we assume that the schedule and win probabilities are independent. We distinguish the structure of competition, which depends primarily on the win probabilities, from the dynamics of a tournament which depend on both the win probabilities and the schedule.

The win probabilities may be conveniently represented using a competition network, $\mathcal{G}_{\rightleftarrows} = (\mathcal{V}, \mathcal{E})$. Assign each competitor a node in the network. Introduce a pair of directed edges between each pair of competitors who could compete with each other. The edge from B to A is assigned the weight p_{AB}. In all that follows we will assume that the tournament is finite, *connected* and *reversible*. That is there are finitely many competitors, for any pair of competitors A B there is a path from A to B and from B to A through $\mathcal{G}_{\rightleftarrows}$ with probability greater than zero, and that $p_{AB} \neq 0$ or 1.

Sometimes it is preferable to simplify the competition network by rounding all weights less than $1/2$ to 0, and all weights greater than $1/2$ to 1. This can be conveniently represented as an unweighted graph $\mathcal{G}_{\rightarrow}$ which contains all directed edges from $\mathcal{G}_{\rightleftarrows}$ with weights greater than a half, and an undirected edge between all pairs with $p_{AB} = 1/2$. This graph represents the expected direction of each competition event, as opposed to the probability

[2]This is distinct from a *complete* tournament in which it must be possible for all pairs to compete.

of each event. Most intransitivity measures focus on this graph (see [97], [117], [104]).

A *ranking* is an ordered list of competitors from best to worst. This can be specified by a rank function R which returns the rank of each competitor. Note that this is distinct from a *rating*, r, which is a function that returns a real number for each competitor [123]. Rankings are often generated by first generating a rating for each competitor, then listing them in decreasing order. Rankings and ratings provide an intuitive description of competition in which some innate competitive ability determines the performance of each competitor against all opponents.

Ranking methods are diverse, and well studied. Famous examples include the Page-rank method used by Google to sort search results [124], the Massey and Colley methods used by the NCAA to rank basketball and football teams [123], and the Elo rating/ranking widely used by chess federations [122, 125]. The rating system produced by the HHD is a kind of log-least squares rating as is frequently used in paired comparison [96, 126, 127]. Examples of least squares rating systems are included in [128, 129, 123, 130, 131, 132]. A survey of least squares rating systems and a comparison to the ratings produced by the HHD is provided in Sections 4.4.1 and 4.5.4.

A competitive network $\mathcal{G}_{\rightleftarrows}$ is consistent with a ranking R if $A > B$ whenever $R(A) < R(B)$. If a competitive network is consistent with a ranking then this ranking is unique and the network is *transitive*. Transitive networks satisfy the intuitive property that if we consider some sequence of competitors with monotonically increasing rank, $A > B > C > D$ then $A > D$. That is, $\mathcal{G}_{\rightarrow}$ contains no cycles, and all the edges in $\mathcal{G}_{\rightarrow}$ point from competitors who have high ranks (low ratings) to competitors with low ranks (high ratings).

If $\mathcal{G}_{\rightarrow}$ contains a cycle, then there exists a sequence of competitors such that $A > B > C > .... > A$, and the tournament is *intransitive*. If a network is intransitive then it is not consistent with any ranking [100]. Speaking broadly, measures of intransitivity either

count the number of intransitive triangles present in $\mathcal{G}_{\rightarrow}$ [97], or measure how far $\mathcal{G}_{\rightarrow}$ is from a nearby transitive network [104]. The Kendall measure [97] counts the number of intransitive triangles in $\mathcal{G}_{\rightarrow}$. This can be done efficiently, however prioritizes triangles over larger loops and does not weight edges equally [133, 104]. The Slater measure of intransitivity is the minimum number of edge directions that need to be reversed in order to transform $\mathcal{G}_{\rightarrow}$ into a transitive network [104]. While conceptually preferable [16], finding the closest transitive network is an NP hard problem [134], [135], [136], [16]. Despite some fast heuristics [116], this limits the application of the Slater measure to small networks. The intransitivity measure associated with the HHD is conceptually analogous to the Slater measure, but can be computed efficiently even for very large networks. Note that transitivity and intransitivity are defined relative to the direction of competition, that is, the *sign* of $p_{AB} - 1/2$, rather than the exact value p_{AB}. In contrast the intransitivity measure associated with the HHD is continuous in the win probabilities, so uses all the information available in $\mathcal{G}_{\rightleftarrows}$. A survey of intransitivity measures and a comparison to the intransitivity measure associated with the HHD is provided in the Section 4.4.

4.4 Survey of Existing Methods

4.4.1 Least Squares Ranking Methods

The most direct method for rating competitors in a tournament is to rate each competitor by their win percentage. That is, if competitor A plays n_A games and wins W_A of them then competitor A is assigned the rating W_A/n_A. The competitors are then ranked in decreasing order by their win percentages [123]. This method is appealingly simple, and is widely used within conferences in professional sports including the NFL, NBA, NHL, and MLB

[128]. That said, rating by win percentages is susceptible to bias because it does not take into account which opponents a competitor frequently faces. If the schedule pits A against strong opponents more often than against weak opponents then the win percentage of team A will underestimate their expected win percentage against an average team. This sort of bias is minimized in sports where each team plays a representative fraction of the conference or league[3], however is a serious problem when the number of competitors is large and the number of games is small [129, 130]. For example, the 117 college football teams each only play 11 games a season, so cannot possible play a representative sample of their league [128].

Biases of this kind motivate rating systems that take into account the strength of schedule. These include the rating system proposed by Massey, Colley, and Keener for ranking college football teams [128, 129, 123, 130].[4] Both the Colley and Massey methods are examples of a least squares rating method. Least squares rating methods attempt to find a rating that is a “best fit" to an edge flow which reflects the win probabilities.

An edge flow is an alternating function f on the directed edges of a graph [16]. Here alternating means that, given a pair of competitors A, B, $f_{AB} = -f_{BA}$. In general the edge flow is chosen to reflect the win probabilities, so that a large positive edge flow from A to B indicates a high probability that A beats B, while a large negative edge flow from A to B indicates a low probability that A beats B. If $f_{AB} = 0$ then the probability A beats B should be 1/2. Examples of edge flows include:

$$f_{AB} = p_{AB} - \frac{1}{2}, \quad f_{AB} = \text{logit}(p_{AB}) = \log\left(\frac{p_{AB}}{1 - p_{AB}}\right). \tag{4.1}$$

[3]Baseball, basketball, professional football, or hockey

[4]Ranking is historically important in college football, where the winner each year was determined by rankings, not victory in a play-off series [128, 130]. Rankings that did not account for strength of schedule could lead to controversy when declaring a champion [129].

The former arises naturally when using methods based on win frequencies [128, 123], while the latter arises naturally from Elo type rating systems [137, 138, 123]. The latter is often referred to as a log-odds, or logit, edge flow.

Given an edge flow we need a method to aggregate the edge flows into a set of ratings that accounts for the strength of different competitor's schedules.

Let r be a rating function that returns an estimate of the competitive ability of each competitor. Since competitive ability does not have an absolute scale we will assume that r is always chosen so that the sum of the ratings equals zero. In order for r to be interpretable the probability that A beats B should be related to the difference in r_A and r_B. A particularly simple choice would be to look for r such that the difference in ratings matches the edge flow [123, 130]:

$$r_A - r_B = f_{AB} \tag{4.2}$$

Then competitor A is rated higher than competitor B when there is a large edge flow from B to A (high probability A beats B).

Notice that, if there exists an r which satisfies equation eq. (4.2) then if $p_{AB} = 1/2$ then $r(A) - r(B) = 0$, so $r(A) = r(B)$. Moreover, if A and C are not directly connected, but are connected through a shared neighbor B, then $r(A) - r(C) = r(A) - r(B) + r(B) - r(C) = f_{AB} + f_{AC}$. To see that this accounts for differences in strength of schedule suppose that A only ever plays B who plays both A and C. Suppose that A and B are equally matched, but C is worse than both. Then B will have the best win record of the three, so would be rated higher than A using win frequency. However, using eq. (4.2), $r(A) = r(B) > r(C)$ so the rating system correctly rates A the same as B, thus accounting for the fact that A has a tougher schedule than B.

Equation eq. (4.2) is a linear equation that maps from a function r on the m competitors, to a function f on each edge of $\mathcal{G}_{\rightleftarrows}$. Let E denote the number of edges in the network. If the network is a tree (contains no loops) then $E = m - 1$. Otherwise $E \geq m$. Since we required that $\sum_{j=1}^{m} r(j) = 0$ the rating function only has $m - 1$ degrees of freedom. It follows that, whenever the competitive network contains a loop, there may not exist any r which satisfies eq. (4.2). Therefore, instead of looking for an r that recovers the edge flow exactly, we look for the rating that most closely recovers the edge flow [131, 132]. [5]

A natural proposal is to minimize the least squares error between $r(A) - r(B)$ and f_{AB} on all edges. This error may be weighted by a set of weights $w_{AB} \geq 0$. Prior information about the ratings may be incorporated by introducing a regularization term $\mathcal{R}(r) \geq 0$ which pushes the final ratings away from ratings where the regularization is large. For example, if $\mathcal{R}(r) = \alpha||r||^2$ for some $\alpha > 0$ then the regularization term helps ensure that none of the ratings are too extreme. A general least squares rating system adopts ratings r which satisfy:

$$r = \text{argmin}_{u \in \mathbb{R}^m : \sum u = 0} \{ \sum_{ij \in \mathcal{E}} w_{ij} \left[(u_i - u_j) - f_{ij} \right]^2 + \mathcal{R}(u) \}. \tag{4.3}$$

The Massey systems sets w_{ij} equal to the number of events observed between i and j, f_{ij} to the point differential, and does not introduce any regularization [123]. If the point differential is modified to account for home field advantage then this is the Stefani system [132]. The Colley system also sets w_{ij} to the number of events observed between

[5] The existence of an exact solution to eq. (4.2) depends on the presence of loops because of the possibility of intransitivity. If there is an intransitive loop in the network then there could be some cycle around which f are all positive. There is no rating that can capture this intransitivity, since if equation eq. (4.2) is satisfied then following a path in $\mathcal{G}_{\rightarrow}$ in the positive direction of an edge flow would lead to a monotonic increase in ratings. This is impossible if there is a cycle around which the edge flows all point in the same direction, since, starting from A, and arriving at a B later in the cycle we would conclude $r(B) > r(A)$, but starting from B and moving in the same direction to A we would conclude $r(A) > r(B)$. Therefore equation eq. (4.2) can not be satisfied exactly if the network is intransitive.

each pair, but sets f_{ij} equal to one half the number of observed wins minus the number of observed losses, divided by the number of games plus two. Note that this is close to the win frequency minus one half. Colley also uses $\mathcal{R}(r) = 2||r||^2$. This suppresses overly large or small ratings. This regularization can be derived from Laplace's rule of succession when estimating the win probability from an observed win frequency [128, 123]. This leads to more conservative ratings.

Alternatively, setting f_{ij} to the log-odds recovers the family of logarithmic least squares rating systems used in the pairwise comparison literature [139, 140, 126, 127]. Setting w_{ij} equal to the number of observed events and using the log-odds produces a system that has been used to rank professional tennis players [96], to rank items in paired comparison studies [126, 127], and to rank the wealth of nations [141]. This method is often used in decision theory when responses to paired comparisons are assumed to be ratios of non-negative quantities [126, 127, 141], or to find an approximate solution to the Bradley-Terry model. [6]

The ratings produced by our application of the HHD are least squares ratings with a logit edge flow. In this chapter the weights are all set to one and no regularization is used as it is assumed that the win probabilities are known. When, as in most empirical settings, the win probabilities are unknown, but estimable from observed win frequencies, then weights and regularization can be chosen based on Bayesian considerations (see Appendix B.1.1).

[6] As an example, a log least squares method proposed by Sismanis, Elo++, won the kaggle chess ratings competition [142]. The kaggle chess rating competition was an open source competition that provided data on 73,000 games among 8,000 competitors. The data was divided into a training data set of 65,000 games, and a test set of the remaining 8,000. Competitors were allowed to train ratings systems on the training set, and then were ranked based on the accuracy of their rating systems when used to predict the outcomes of the test set. The Elo++ method uses least squares ranking with a logit edge flow and a bias associated with a player playing white or black, and with the weights chosen to emphasize recent games. Like Colley, Sismanis regularized the least squares rating problem. Unlike Colley, whose regularization penalized large ratings, Sismanis' regularization term penalized large differences between the rating of a competitor and their neighbors. This choice of regularization was motivated by the observation that most chess players primarily play opponents of similar ability [142].

4.4.2 Measures of Intransitivity

Existing intransitivity measures can be broadly broken into two categories: measures based on triangle census, and measures of the distance to nearby transitive networks.

Triangle Census:

Consider a triangle consisting of three competitors A, B, C. Assume that these competitors are all connected in $\mathcal{G}_{\rightleftarrows}$. Then either the triangle is intransitive (the directed edges in $\mathcal{G}_{\rightarrow}$ point around the cycle), or it is transitive.

Kendall [97] proposed measuring the transitivity, K, of a network by counting the total number of all triangles in $\mathcal{G}_{\rightarrow}$ that are intransitive, normalizing this count by the maximum possible, and then subtracting this ratio from 1. This measure was originally designed for pairwise comparison of objects, where it was assumed that all pairs could be compared with each other [97]. That is, the method was developed for complete tournaments.

For a complete tournament the total number of intransitive triangles can be computed analytically without counting over all triangles by computing the variance in the in-degree of the nodes of $\mathcal{G}_{\rightarrow}$ [97]. This is the principle advantage of Kendall's measure [95].

Let k denote the number of intransitive triangles in $\mathcal{G}_{\rightarrow}$ and let k_{max} be the maximum possible. Kendall defined his measure of transitivity (or "consistency") to be $K = 1 - \frac{k}{k_{max}}$. Therefore the associated measure of intransitivity is:

$$\text{Int}_{\text{K}}(\mathcal{G}_{\rightarrow}) = \frac{k}{k_{max}}. \tag{4.4}$$

This means that, if the tournament is complete then the Kendall measure $\text{Int}_{\text{K}}(\mathcal{G}_{\rightarrow})$ can be computed analytically by computing the variance in the in-degree of each node. In general, the larger this variance the more transitive the tournament.

Landau defined a related measure, h which is simply the variance in the in-degree of each node, scaled by the maximum possible variance [117]. This is equivalent to Kendall's K if the number of competitors is odd [133, 95]. As written this is a transitivity measure since it increases as the network becomes more transitive. We will refer to $\text{Int}_{\text{L}}(\mathcal{G}_{\rightarrow}) = 1-h$ as Landau's intransitivity measure. This measure was rediscovered by Laird [94], and was used to show that intransitive competition between species promoted coexistence.[7]

Both the Kendall measure and the Landau measure are restricted to complete tournaments. More general approaches are needed since most data sets are far from complete [95, 16].

If the underlying tournament is incomplete then it is still possible to perform a triangle census explicitly by checking every triangle. This idea was proposed by both Shizuka [95] and de Vries [143]. Shizuka's measure is defined as the proportion of all triads that are intransitive:

$$\text{Int}_{\text{Sh}}(\mathcal{G}_{\rightarrow}) = \frac{\text{number of intransitive triangles}}{\text{number of triangles}}. \tag{4.5}$$

Here the normalization is done with respect to all triangles, since it is easier to find the total number of triangles than the maximum number of intransitive triangles [95]. If the direction of all edges are chosen randomly then on average a quarter of all triangles are intransitive [95, 144]. Therefore, observing $\text{Int}_{\text{Sh}}(\mathcal{G}_{\rightarrow}) < 0.25$ is an indication that the network is more transitive than would be expected if all pairwise relations were chosen randomly. Most empirical networks exhibit $\text{Int}_{\text{Sh}}(\mathcal{G}_{\rightarrow}) < 0.25$ [95].

Shizuka's method differs from other popular methods which start by either randomly filling in missing edges [143], or by replacing missing edges with a pair of edges each

[7] Laird et al normalize the measure slightly differently. They normalize by the difference between the maximum possible variance and the minimum possible variance. The minimum possible variance is nonzero if m even.

weighted by 1/2 [133]. Once all missing edges have been filled either the Kendall or Landau metrics can be used. These methods should be avoided if the underlying network is not complete (missing edges due to structure not lack of data), or if there is correlation structure in the orientation of the edges that is not accounted for by randomly filling in the missing edges. It has been observed that these imputation procedures typically overestimate the degree of intransitivity [95]. Using the metric of intransitivity associated with the HHD we will explicitly show that adding edges, especially adding edges without incorporating correlation structure, typically increases the expected degree of intransitivity. For this reason it is important to develop measures of intransitivity that, like Shizuka's measure, do not require randomly filling in missing data.

Distance to Transitive Network:

An alternative family of intransitivity measures follow from the measure proposed by Slater [104]. Slater proposed measuring the intransitivity of a competitive network by counting the minimum number of competitive reversals (flipped edges in $\mathcal{G}_{\rightarrow}$) needed to reach a transitive network [104]. Unlike the Kendall measure, Slater's index of intransitivity does not treat triangles as fundamental units. Note that, in a tournament that is not complete there may be intransitive cycles of length greater than 3, but no intransitive cycles of length 3.[8] The Kendall measure also puts more weight on the orientation of edges that appear in many triangles than in few triangles [133, 104]. Slater's measure weights edges equally and does not emphasize triangles over loops of other sizes [104]. For this reason the Slater measure is sometimes considered conceptually preferable [16]. This same measure was rediscovered by Petraitis, in the context of complete tournaments. Petraitis normalized

[8]For example, if the network consists of many large loops but no triangles these loops may well be intransitive, but that intransitivity cannot be measured with a triangle census.

the measure by the maximum number of possible reversals needed to reach the closest transitive tournament [100].

The Slater measure is defined by a choice of metric on graphs. If $\mathcal{G}_{\rightarrow}$ and $\mathcal{G}'_{\rightarrow}$ are two different directed graphs that differ only in the direction of their edges then the Kendall τ distance (or bubble sort distance) is the total number of edges in the two directed graphs that point in different directions. We will denote this $\tau(\mathcal{G}, \mathcal{G}')$. Let $\mathcal{T}$ denote the family of all transitive graphs. Then the Slater index of intransitivity can be written:

$$\text{Int}_{\text{Sl}}(\mathcal{G}_{\rightarrow}) = \min_{\mathcal{G}'_{\rightarrow} \in \mathcal{T}} \tau(\mathcal{G}_{\rightarrow}, \mathcal{G}'_{\rightarrow}). \tag{4.6}$$

A transitive network that minimizes the distance to the original network is Kemeny optimal. Note that any transitive network is consistent with a unique ranking of the competitors. A Kemeny optimal ranking is a ranking which generates a competition network as close as possible to the original network. In essence this is a ranking that leads to the fewest observed upsets [129]. Note that there may be more than one transitive $\mathcal{G}'$ that is a distance $\text{Int}_{\text{Sl}}(\mathcal{G}_{\rightarrow})$ from $\mathcal{G}$, hence more than one Kemeny optimal rating.

This method can be generalized by assigning each edge in $\mathcal{G}_{\rightleftarrows}$ a weight equal to p_{AB} or the number of observed times A beat B, then letting τ be the sum of the weights on the edges of $\mathcal{G}_{\rightleftarrows}$ that do not appear in $\mathcal{G}_{\rightarrow}$ (the total probability or number of upsets). The associated closest transitive network gives the Kemeny optimal ranking. Young showed that this ranking is a maximum likelihood estimator of a true ranking if a true ranking exists and there is a fixed upset probability [145], [146].

Finding Kemeny optima is known to be a NP complete problem [134], [135], [136], [16], as it is equivalent to solving the minimum feedback arc-set problem [115]. While some fast heuristics exist [116], in practice the Slater measure is often implemented using

a brute force search over all possible $m!$ rankings as in [94].[9] As a result Slater's measure is used less widely than Kendall's.

An alternative to the Slater index that is easier to use in some contexts was proposed by Ulrich [114]. Instead of looking for a Kemeny optimal ranking, Ulrich proposed measuring the distance between the original network, and a network generated by a relevant, easy to compute ranking. For example, if the tournament can be cast as a Markov chain then the steady state distribution of the Markov chain can be considered a rating of each competitor, and the associated ranking can be used to generate a transitive directed network $\mathcal{G}'_{\rightarrow}$. Then the distance between $\mathcal{G}_{\rightarrow}$ and $\mathcal{G}'_{\rightarrow}$ can be used as a measure of intransitivity This is numerically much easier because the steady state is an eigenvector of the transition matrix, so the problem of finding the transitive network to compare to reduces to an eigenvector problem.[10] This approach was adopted by [113] in a large scale study of the degree of intransitivity present in grassland ecosystems.

A summary of the methods discussed thus far is included below. Note that some of the methods are restricted to complete graphs because the appropriate normalization constant is only known for complete graphs. If these measures were not normalized then they could easily generalize to other network topologies.

The intransitivity measure associated the HHD is attractive since it retains the conceptual clarity of Slater's approach, while remaining computationally tractable, and in the case of complete graphs, is analytically tractable. Like Slater's measure the intransitivity measure associated with the HHD measures the distance to a closest transitive network. The two differ in the choice of metric used to define "closest". Slater's metric depends on

[9] Note that, if the graph is sufficiently sparse, then it is more efficient to perform a brute force search over all 2^E possible directed graphs.

[10] As originally proposed this measure is normalized by the maximum number of possible flips. This limits the application to complete graphs where the appropriate normalization constant is known as a function of m.

Source	Symbol	Type	Method	Application	Computation
Kendall [97]	Int_K	Triangle Census	$\frac{\text{number intransitive } \triangle}{\text{max number}}$	Complete Only	Analytic
Landau/Laird [117]	Int_L	Triangle Census	$1 - \frac{\text{variance in-degree}}{\text{max variance}}$	Complete Only	Analytic
de Vries [143]	Int_dV	Triangle Census	mean Int_L	Missing Data	Random Sampling
Shizuka [95]	Int_Sh	Triangle Census	fraction $\triangle$ intransitive	All Networks	Triangle Search
Slater [104]	Int_Sl	Nearest Transitive	fewest flips to transitive	All Networks	NP Hard
Petraitis [100]	Int_P	Nearest Transitive	$\frac{\text{fewest flips to transitive}}{\text{max flips}}$	Complete Only	NP Hard
Ulrich [114]	Int_U	Nearest Transitive	$\frac{\text{fewest flips to steady state}}{\text{max flips}}$	Complete Only	Eigenvector

Table 4.1: Summary of the six intransitivity measures discussed.

the number of competitive reversals, which is not continuous in the win probabilities. In contrast the metric used by the HHD is continuous in the win probabilities, and leads to a trivial optimization problem with a unique solution. This optimization problem does not change when the network structure changes, and, as a consequence the intransitivity measure produced by the HHD allows for completely arbitrary network structure. Therefore, using the HHD does not require filling in missing edges, or introducing a normalization factor that is difficult to compute.

The intransitivity measure associated with the HHD also differs from all of the metrics presented here in that it is continuous in the win probabilities p.[11] This is an important distinction since it allows the measure to distinguish between a loop $A \to B \to C \to A$ with win probabilities 0.51 on each edge from the same loop with win probabilities 0.99 on each edge. This leads to a more nuanced view of intransitivity, which is detailed in the following section.

4.5 The Network HHD

The Network Helmholtz-Hodge Decomposition (HHD) can be derived by defining two special classes of tournaments. These parallel the two classes of games defined in [18].

4.5.1 Arbitrage Free and Favorite Free Tournaments

Arbitrage Free Tournaments (Perfectly Transitive)

A currency market is said to be *arbitrage free* if it is impossible to make money by exchanging currencies in a cyclic fashion [16]. By analogy we define an *arbitrage free tournament*

[11] Petraitis [100] proposed an extension of his measure to the continuous case by changing the metric between directed graphs to a metric on weighted directed graphs, where the metric is set to the sum of the absolute value of the difference in weights across all edges.

to be a tournament for which it is impossible to expect to make money by betting on cyclic sequences of events. Specifically, a tournament is arbitrage free if, for any cyclic sequence of competitors $\mathcal{C} = \{i_1, i_2, \ldots, .i_n, i_{n+1} = i_1\}$, a sequence of wins where i_j loses to i_{j+1} (i_1 loses to i_2 loses to i_3 and so on) is equally likely as a sequence of wins where i_j beats i_{j+1} (i_1 beats i_2 who beats i_3 and so on). This requires that the win probabilities satisfy a cycle condition.

Cycle Condition: A tournament is arbitrage free if and only if, for every cycle $\mathcal{C} = \{i_1, i_2, \ldots, i_n, i_{n+1} = i_1\}$, the win probabilities satisfy:

$$p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_n i_1} = p_{i_1 i_n} \cdots p_{i_3 i_2} p_{i_2 i_1}. \tag{4.7}$$

The cycle condition can be expressed more simply by dividing the right hand side across to the left hand side and then taking a logarithm. This gives the equivalent condition:

$$\sum_{j=1}^{n} f_{i_j i_{j+1}} = 0 \tag{4.8}$$

where the f_{ij} is the log-odds that competitor i beats competitor j:

$$f_{ij} = \text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right). \tag{4.9}$$

Therefore the cycle condition is satisfied if and only if the sum of f around any cycle is zero. The log-odds, f, are an example of an edge flow, an alternating function, $f_{ij} = -f_{ji}$, on the edges [16].

Lemma 15 (Arbitrage Free Tournaments are Transitive)**.** *A tournament is arbitrage free if and only if its win probabilities are consistent with a unique set of ratings r that satisfy*

$p_{ij} = \text{logistic}(r_i - r_j)$ constrained to $\sum_i r_i = 0$ [12]. Moreover if a tournament is arbitrage free then it is transitive.

Proof. Suppose that a tournament is arbitrage free. Then it must satisfy the cycle condition. This implies that the sum of f around any cycle is zero. It follows that, for any pair of endpoints A, B, the value of the sum of f over a path connecting A to B is path independent.

To recover the associated ratings, pick an arbitrary spanning tree of the network and an arbitrary starting competitor A.[13] Then let u_B equal the sum of f over the path connecting A to B in the tree. Finally let $r_B = u_B - \frac{1}{m}\sum_i u_i$. Then, by construction, $\sum_i r_i = 0$. It remains to show that $r_i - r_j = f_{ij}$ for all connected pairs i, j. By construction, this must be true for all i, j that are connected through an edge in the spanning tree. Consider an edge not in the spanning tree (a chord) connecting i and j. Let $i_1 = A, i_2, \ldots, i_l = i$ and $j_1 = A, j_2, \ldots, j_k = j$ be the paths from A to i and j through the spanning tree. Then $r_i - r_j = u_i - u_j = \sum_{n=1}^{l-1} f_{i_{n+1}i_n} - \sum_{n=1}^{k-1} f_{j_{n+1}j_n} = \sum_{n=k}^{2} f_{j_{n-1}j_n} + \sum_{n=1}^{l-1} f_{i_{n+1}i_n}$ which is the sum over the path from j to A then from A to i. If the chord was added to the path then this would complete a loop from j to A to i back to j (see Figure 4.1). By assumption the sum of f around any loop is zero, so $r_i - r_j + f_{ji} = r_i - r_j - f_{ij} = 0$, or, $r_i - r_j = f_{ij}$. Therefore, if a tournament is arbitrage free then there exist a set of ratings r such that $r_i - r_j = f_{ij}$. Since $f_{ij} = \text{logit}(p_{ij})$ this implies $p_{ij} = \text{logistic}(r_i - r_j)$. These ratings are unique since the sum of f is path independent, hence the ratings generated by the spanning tree construction are independent of the choice of tree.

Suppose that $p_{ij} = \text{logistic}(r_i - r_j)$. Then $f_{ij} = r_i - r_j$ for all connected i, j. This means that, given a path $i_1, i_2, \ldots, i_n$ the sum $f_{i_2 i_1} + f_{i_3 i_2} + \ldots f_{i_n i_{n-1}} = r_{i_n} - r_{i_1}$ as the

[13] A spanning tree is a subgraph of the network that contains no loops, includes all competitors, and is connected.

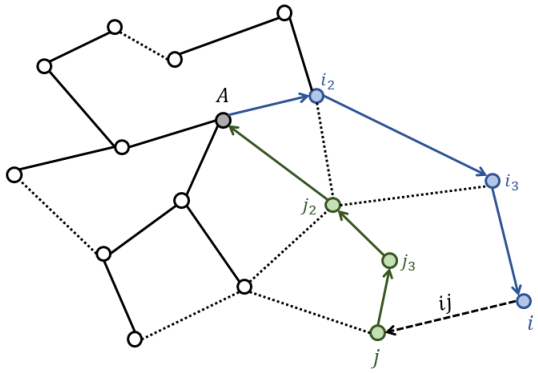

Figure 4.1: The spanning tree construction for recovering the ratings for an arbitrage-free tournament. The tree is shown with solid lines, and the chords with dotted lines. The root of the tree, A is marked in grey. Two vertices, i and j connected by a chord ij, are shown in blue and green respectively. The sequence of nodes leading from A to i and j are labelled. Then, by the cycle condition, the sum around the loop marked with arrows is zero, hence $f_{ij} = r_i - r_j$.

sum is telescoping. If the path is a loop then $i_n = i_1$ so the sum equals zero. This means that f satisfies the cycle condition, so the tournament is arbitrage free.

Suppose the tournament is arbitrage free. Then $p_{ij} = \text{logistic}(r_i - r_j)$ for a unique set of ratings r. This means that $p_{ij} > 1/2$ if and only if $r_i > r_j$. It follows that $A > B$ if and only if $r_A > r_B$, so the win probabilities are consistent with the ranking induced by the ratings r. This means that the tournament is transitive. □

Theorem 15 shows that arbitrage free tournaments are the only tournaments which exactly match the logistic rating model $p_{ij} = \text{logistic}(r_i - r_j)$. This is the model assumed by the Elo rating system [137, 138, 123].[14]

Arbitrage free tournaments are also the only tournaments which match the Bradley-

[14]The Elo rating system was originally proposed to rate chess players, but is also used to rank Sumo wrestlers [125], English league football teams [138] and international football teams. In the latter example the Elo method was the most predictive out of all methods tested [147]. The Women's World Cup uses a variant on the Elo method [147].

Terry model:[15] $p_{ij} = q_i/(q_i + q_j)$ where $q_i \geq 0$ for all i [139, 140]. If a network is arbitrage free, then from setting $q_i = \exp(r_i)$ it follows that $p_{ij} = q_i/(q_i + q_j)$. Alternatively, if the tournament satisfies the Bradley-Terry model, then setting $r_i = \log(q_i)$ produces a rating which satisfies $p_{ij} = \text{logistic}(r_i - r_j)$, so the network must be arbitrage free. The values, q, which appear in the Bradley-Terry model are widely used as ratings.

Since arbitrage free networks are a special class of transitive networks, we will refer to these networks as "perfectly" transitive. Note that a perfectly transitive network must satisfy the cycle condition, which is a requirement on the values of p rather than simply the sign of $p - 1/2$. Hence, while all perfectly transitive networks are transitive, not all transitive networks are perfectly transitive. For example, if $p_{AB} = 0.99$, $p_{BC} = 0.99$, and $p_{AC} = 0.51$ then the tournament is transitive, even though p_{AC} is much smaller than might be expected given p_{AB} and p_{BC}. This tournament is not perfectly transitive since it does not satisfy the cycle condition.

Favorite Free Tournaments (Perfectly Cyclic)

In contrast to arbitrage free tournaments, we define a *favorite free tournament* to be a tournament for which it is impossible to make money on average by betting on a favorite competitor over his or her neighbors. Specifically, we require that in a favorite free tournament A is equally likely to beat all of their neighbors, as to lose to all of their neighbors. This leads to a neighborhood condition.

Neighborhood Condition: A tournament is favorite free if and only if, for every com-

[15] The Bradley-Terry model is widely used in pairwise comparison and to rank competitors in tournaments. Examples include professional tennis [148], Cape dwarf chameleons [118] and northern elephant seals [119]. Bradley-Terry models accounting for surface type, and discounting old games, have been shown to be effective in predicting the outcome of ATP tennis tournaments, consistently outperforming standard rankings [148]. In a meta-study of predictive models the Bradley-Terry model had moderate predictive accuracy when compared to regression based methods, but was generally outperformed by Elo based methods which were the most accurate of all methods tested [149].

petitor i with neighborhood $\mathcal{N}(i)$, the win probabilities satisfy:

$$\prod_{j \in \mathcal{N}(i)} p_{ij} = \prod_{j \in \mathcal{N}(i)} p_{ji}. \tag{4.10}$$

Like the cycle condition, the neighborhood condition can be written directly as a condition on the log-odds edge flow f. Dividing across by the left hand side and taking a logarithm we see that a tournament satisfies the neighborhood condition if and only if the sum of f_{ij} over the neighborhood of i is zero for all competitors i:

$$\sum_{j \in \mathcal{N}(i)} f_{ij} = 0. \tag{4.11}$$

If the neighborhood condition is satisfied then it can be extended to all sets of competitors. Let S be a set of competitors and let $\mathcal{N}(S)$ be the set of all competitors not in S who neighbor S. Then the neighborhood condition implies:

$$\sum_{j \in \mathcal{N}(S), i \in S} f_{ij} = 0. \tag{4.12}$$

This identity follows from the discrete divergence theorem, which states that the sum of f over the neighborhood of S equals the sum of the divergence of every competitor in S. If i and j are both in S then the sum over the neighborhood of i contributes f_{ij}, and the sum over the neighborhood of j contributes $f_{ji} = -f_{ij}$. Therefore all the internal edges cancel in the sum. So $\sum_{j \in \mathcal{N}(S), i \in S} f_{ij} = \sum_{i \in S} \sum_{j \in \mathcal{N}(i)} f_{ij} = \sum_{i \in S} 0 = 0$.

The cycle condition defined a special subset of transitive tournaments. The neighborhood condition also defines a special class that can be seen as a subset of a larger class - the class of *cylic* tournaments.

We define a *cyclic tournament* to be a tournament such that, if there is a path from A to

B in $\mathcal{G}_{\rightarrow}$, then there must be a path back from B to A in $\mathcal{G}_{\rightarrow}$.

Lemma 16 (Favorite Free Tournaments are Cyclic)**.** *A favorite free tournament is cyclic, and is never transitive unless $p_{ij} = 1/2$ for all connected ij.*

Proof. Suppose that a given tournament is favorite free. Then $\sum_{j \in \mathcal{N}_i} f_{ij} = 0$ for all i. This leaves two distinct possibilities, either $f_{ij} = 0$ for all $j \in \mathcal{N}(i)$, or there is some j such that $f_{ij} \neq 0$. The former case requires $p_{ij} = 1/2$ for all $j \in \mathcal{N}(i)$. We will refer to this case as the *neutral* case. If the neighborhood of i is not neutral then $f_{ij} \neq 0$ for some $j \in \mathcal{N}(i)$. Since the sum over all j is zero this means that there must be at least one other edge ik such that $\text{sign}(f_{ij}) = -\text{sign}(f_{ik})$. This means that, if there is an edge into competitor i in $\mathcal{G}_{\rightarrow}$ there must also be at least one edge out of i in $\mathcal{G}_{\rightarrow}$ (recall that if $p_{ij} = 1/2$ then there are a pair of edges between i and j, one from i to j and one from j to i).

Since the neighborhood condition can be extended from the neighborhood of competitors to the neighborhood of sets this property can also be extended to sets. That is, if there is an edge into the set S in $\mathcal{G}_{\rightarrow}$ then there must also be an edge out of the set S in $\mathcal{G}_{\rightarrow}$.

Now suppose that there is a path from A to B in $\mathcal{G}_{\rightarrow}$. It remains to construct a path back to A.

Define the nested sets $S_0(B), S_1(B), \ldots,$, where $S_d(B)$ is the set of all nodes that can be reached from B with a path in $\mathcal{G}_{\rightarrow}$ of length less than or equal to d. Now since there is a path from A to B in $\mathcal{G}_{\rightarrow}$ there is an edge in $\mathcal{G}_{\rightarrow}$ arriving at $\{B\} = S_0(B)$. Thus there is a path from A to all competitors in $S_1(B)$. Now there are two possibilities, either A is in $S_1(B)$, or A is not in $S_1(B)$. If A is in $S_1(B)$ then we are done. If not, then there is an edge entering $S_1(B)$ in $\mathcal{G}_{\rightarrow}$ since there is a path from $A \notin S_1(B)$ to $B \in S_1(B)$. Then the neighborhood condition implies that there is an edge out of $S_1(B)$, which means that $S_2(B) \neq S_1(B)$. Now the logic repeats. Either A is in $S_2(B)$, in which case we are done,

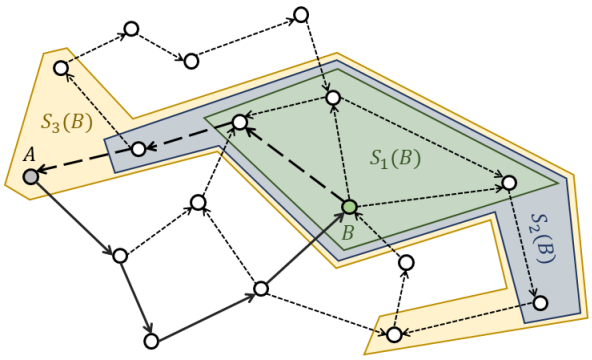

Figure 4.2: A favorite free tournament must be a cyclic tournament. The arrows represent the direction of competition. If the network is favorite free, then, if there is an edge pointing into a set, there must be an edge pointing out of it. A path from A to B is shown in black. Then the sets $S_1(B), S_2(B), S_3(B)$ are shown as shaded polygons. These contain all competitors distance 1, 2, and 3 (respectively) from B. These sets continue to expand until they include A, hence there is a path from B to A.

or it is not. If it is not then there must be an edge entering $S_2(B)$ so there must be an edge leaving $S_2(B)$ so $S_3(B) \neq S_2(B)$. This means that, as long as $A \notin S_d(B)$ there is a larger set $S_{d+1}(B) \neq S_d(B)$ which can be reached from B. Since we assumed that there are finitely many competitors this can only continue until A is contained in $S_d(B)$ for some B. This proof technique is illustrated in Figure 4.2.

Suppose that the tournament is transitive, favorite free, and not neutral. Since it isn't neutral there must be at least one pair ij such that $p_{ij} > 1/2$. This means that $r_i > r_j$ and there is an edge from j to i in $\mathcal{G}_{\rightarrow}$. But, if the tournament is favorite free then there must be some other path from i back to j in $\mathcal{G}_{\rightarrow}$. This means that $r_j > r_i$ since there is a path in $\mathcal{G}_{\rightarrow}$ from j to i. This is clearly a contradiction. This implies that a cyclic tournament is not transitive unless it is neutral: $p_{ij} = 1/2$ for all ij.[16] □

So, just as the cycle condition (no tendency to cycle) implied transitivity, the neighbor-

[16]This shows that the two classes of tournaments are distinct, as their only overlap is the neutral case. Note that a neutral tournament is considered transitive since it can be consistently ranked - all competitors should be ranked the same.

hood condition, (no favorites) implies that the network is cyclic, and is only transitive if it is also completely neutral. As before, whether a tournament is cyclic or not depends on the sign of $p_{ij} - 1/2$, while the neighborhood condition is a condition on the values of p_{ij}. This motivates the definition: a tournament is *perfectly cyclic* if and only if it is favorite free. As before, all perfectly cyclic tournaments are cyclic, but not all cyclic tournaments are perfectly cyclic.

Note that, unlike perfectly transitive tournaments where f is determined by a set of ratings r, we are not currently equipped to relate the edge flow of a favorite free tournament to a lower dimensional representation. In Section 4.5.2 we will show that a favorite free tournament has edge flows f which can always be represented as a sum of cyclic intensities (or vorticities) on a set of loops. This result will parallel the conclusions of theorem 15.

4.5.2 The Discrete HHD

Given these two classes of tournaments it is natural to ask: can a generic tournament be decomposed into a perfectly transitive (arbitrage free) part and a perfectly cyclic (favorite free) part? We answer in the affirmative. This is the Helmholtz-Hodge decomposition.

Operators

In order to define the decomposition succinctly it is helpful to have a pair of operators analogous to the gradient and curl operators in the continuum. We simplify the topological presentation in [16] by expressing the decomposition entirely through linear algebra.

First, we define the edge space $\mathbb{R}^E$, where E is the number of pairs i, j who could compete. Index each pair so that edge k has endpoints (competitors) $i(k), j(k)$. Note that this requires assigning each edge an arbitrary start and endpoint so that positive f indicates motion from the start to the end, while negative f indicates motion from the end to the start.

This is simply a sign convention.

Let the *discrete gradient* operator G be the matrix which maps from $\mathbb{R}^m$ to $\mathbb{R}^E$ by setting:

$$[Gu]_k = u_{i(k)} - u_{j(k)}. \tag{4.13}$$

Notice that if r is a rating function on the nodes, then attempting to find r such that $r_i - r_j = f_{ij}$ is equivalent to looking for r such that $Gr = f$. Since any arbitrage free tournament admits a unique rating r such that $Gr = f$ it follows that the space of perfectly transitive networks is equivalent to the space of tournaments with edge flow f in the range of the gradient. Assuming that the tournament is connected, the gradient has a one-dimensional nullspace parallel to the vector $[1; 1; ...1]$. It follows that $G(r + c) = Gr$ if c is some constant. This motivates the constraint $\sum_i r_i = 0$ used throughout, since the edge flow only determines the size of differences in ratings, not the actual ratings.

The gradient transpose, $G^\intercal$ is the discrete divergence operator. The divergence maps from the space of edges to the space of nodes (competitors) such that:

$$[G^\intercal f]_i = \sum_{\mathcal{N}(i)} f_{ij}. \tag{4.14}$$

The neighborhood condition eq. (4.11) is equivalent to requiring that $G^\intercal f = 0$. That is, the space of favorite free tournaments is equivalent to the space of tournaments with edge flow f in the null space of the divergence. Note that, like the divergence operator in the continuum, the discrete divergence obeys the divergence theorem (the sum of the divergence on the neighborhood of each competitor in a set is the same as the sum of the edge flow into the set).

In order to build a parallel description for perfectly cyclic tournaments, we need a space of loops. First define the sum of two cycles $\mathcal{C}_1$, $\mathcal{C}_2$ to be all edges included in either $\mathcal{C}_1$ or

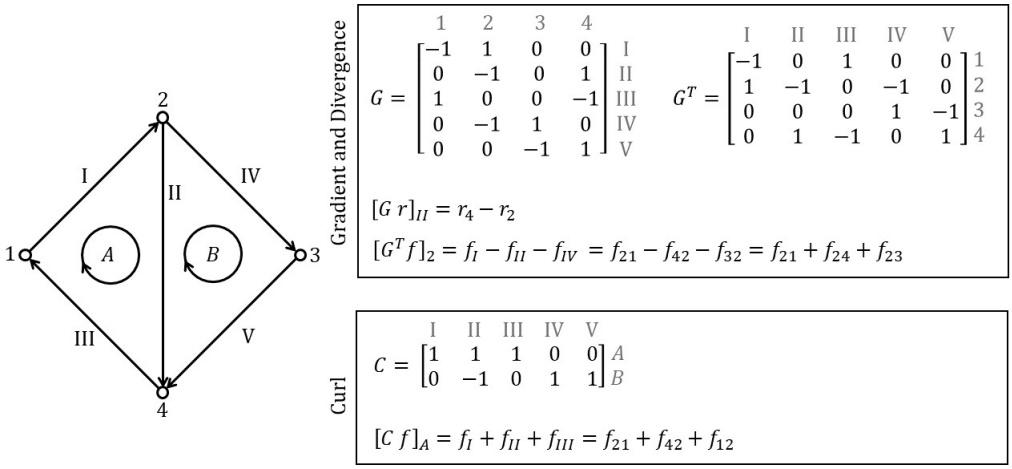

Figure 4.3: The gradient, divergence, and curl for an example network.

$\mathcal{C}_2$ but not both. Equipped with this addition operation, the space of cycles is a vector space, which can be represented with a cycle basis. A *cycle basis* is a collection of linearly independent cycles $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_L$ such that any other cycle $\mathcal{C}$ can be expressed as a linear combination of cycles in the loop basis [150].

Any connected graph admits a cycle basis. A constructive method for finding a cycle basis follows. First, pick a spanning tree of the network. Then the spanning tree includes $m - 1$ edges, and $E - (m - 1)$ edges are left out. These are the *chords*. By construction, the tree does not contain any loops. If one chord is added to the tree then the network contains exactly one cycle. Note that no two chords can produce the same cycle, and that the set of cycles produced by adding the chords to the spanning tree is necessarily linearly independent since no chord appears in more than two of these cycles. Therefore, if we enumerate the chords from $1, 2, \ldots, L = E - m + 1$ then the set of cycles $\mathcal{C}_1, \ldots, \mathcal{C}_L$ associated with each chord is a cycle basis. A basis generated by a spanning tree is a *fundamental cycle basis* [34, 150]. This basis is not unique, since there are often many different possible spanning trees, moreover not all cycle bases need be constructed via a

spanning tree.

Next define the cycle space $\mathbb{R}^L$ to be the space of real vectors with one entry for each cycle in a chosen cycle basis. The dimension of the cycle space $L = E - m + 1$ is the *cyclomatic number* of the network [34, 150]. Then we define the *discrete curl* operator to be the matrix which maps from $\mathbb{R}^E$ to $\mathbb{R}^L$ (edges to cycles) by summing f around each loop. That is, if the set of edges $\{k_1, k_2, \ldots, k_{n_l}\} = \mathcal{C}_l$ then:

$$[Cf]_l = \sum_{h=1}^{n_l} f_{i(k_h)j(k_h)}. \tag{4.15}$$

Note that in order to perform this sum, each loop must be assigned an arbitrary direction of traversal. This is simply a sign convention.

In general we will only consider curl operators that are defined with respect to cycle bases such that there exists an invertible $L \times L$ matrix T for which $TC = \tilde{C}$, where $\tilde{C}$ is the curl operator defined with respect to a fundamental cycle basis.

The curl is analogous to the curl in continuous space, which is a path integral over infinitesimally small loops. Note that the discrete curl defined in this way is more general than the discrete curl defined in [18] or [16]. Jiang and Candogan restrict the curl operator to only act on connected cliques of three nodes (triangles), and then introduce additional operators to account for cliques containing more nodes. This construction can lead to unintuitive conclusions. For example, if $p_{AB} = p_{BC} = p_{CD} = p_{DA} = 0.99$ then there is clearly a cyclic tendency in the competition, but if the curl is restricted to only act on triangles, then the curl of this graph is zero. Here we extend the curl to act on loops of arbitrary length since, like [104], we do not see a fundamental distinction between cyclic structure on triangles and cyclic structure on larger loops. If desired, we could partition the curl operator into blocks, each according to loops of a fixed length, and treat each block as

the curl operator restricted to loops of a given size.

The operators for an example network are provided in Figure 4.3.

Lemma 17 (Orthogonality of Operators)**.** *The curl C and the gradient G are orthogonal, regardless of the choice of cycle basis.*

Proof. Consider the product CGu for some arbitrary vector $u \in \mathcal{R}^m$. The product Gu produces an edge flow, so the product CGu produces a vector whose entries are the sum of that edge flow around a set of loops. Consider an arbitrary path $i_1, i_2, \ldots, i_n$. Then the sum of Gu over the path is $(u_{i_2} - u_{i_1}) + (u_{i_3} - u_{i_2}) + ...(u_{i_n} - u_{i_{n-1}}) = u_{i_n} - u_{i_1}$. Therefore, if the path is a loop, $i_n = i_1$ so the sum is zero. It follows that $CGu = 0$ for all $u \in \mathbb{R}^n$ so:

$$CG = 0, \quad G^\intercal C^\intercal = 0 \tag{4.16}$$

where the second equation follows trivially by transposing the first equation.[17] □

Lemma 18. *If C is a discrete curl operator then if $Cf = 0$, there exists a set of ratings r such that $Gr = f$.*

Proof. This is a direct consequence of theorem 15. If C is a curl operator, then there exists an invertible transform T such that $C = T\tilde{C}$ where $\tilde{C}$ is the curl operator with respect to some fundamental cycle basis. Then $Cf = T\tilde{C}f = 0$ if and only if $\tilde{C}f = 0$. Since $\tilde{C}$ is defined with respect to a fundamental cycle basis, $\tilde{C}$ is defined with respect to a spanning tree $\mathcal{T}$ which generates the cycle basis. Requiring that $\tilde{C}f = 0$ is equivalent to requiring that the sum of f around every loop formed by adding one chord into the tree is zero.

[17] Note that the product GC has no meaning in our framework. Even if the range of C and domain of G were of compatible dimension, the product has no natural interpretation since C maps to loops and G acts on nodes.

This condition is sufficient to reconstruct r such that $Gr = f$ using the spanning tree construction given in the proof of Theorem 15, where the chosen tree is $\mathcal{T}$. □

theorem 17 and theorem 18 establish that any f in the range of the gradient is in the nullspace of the curl, and any f in the nullspace of C is in the range of the gradient. That is, if $f = Gr$ then $Cf = 0$ and if $Cf = 0$ then $f = Gr$ for some rating r. Therefore the range of the gradient is the nullspace of the curl. The equivalence of these two spaces and the orthogonality of the operators allows us to decompose f into unique perfectly transitive and perfectly cyclic components. This is the HHD.

The Discrete Helmholtz-Hodge Decomposition

We are now equipped to prove that every edge flow can be represented as the sum of a perfectly transitive (arbitrage free), and perfectly cyclic (favorite free) edge flow - thus any tournament can be represented as a unique combination of a perfectly transitive and perfectly cyclic tournament. Similar proofs are provided in [18] and [16] .

Theorem 19 (The HHD)**.** *Any $f \in \mathbb{R}^E$ can be decomposed such that:*

$$f = f_t + f_c \tag{4.17}$$

where f_t is arbitrage free (perfectly transitive) and f_c is favorite free (perfectly cyclic):

$$Cf_t = 0, \quad G^\intercal f_c = 0. \tag{4.18}$$

and both are unique. In addition, there exists a unique rating r satisfying $\sum_i r_i = 0$ such that $f_t = Gr$ and for any choice of cycle basis there exists a unique vorticity $v \in \mathbb{R}^L$ such

that $f_c = C^\intercal v$. Thus the original edge flow f can be uniquely decomposed:

$$f = Gr + C^\intercal v. \tag{4.19}$$

Proof. By the fundamental theorem of linear algebra (Fredholm alternative):

$$\mathbb{R}^E = \text{null}(G^\intercal) \oplus \text{range}(G). \tag{4.20}$$

theorem 17 and theorem 18 guarantee that $\text{range}(G) = \text{null}(C)$, so:

$$\mathbb{R}^E = \text{null}(G^\intercal) \oplus \text{null}(C). \tag{4.21}$$

This establishes equation eq. (4.17), where f_t is the orthogonal projection of f onto $\text{null}(C)$ and f_c is the orthogonal projection of f onto $\text{null}(G^\intercal)$.

To prove that the arbitrage free and favorite free fields can be expressed using ratings and vorticities, write:

$$\mathbb{R}^E = \text{null}(C) \oplus \text{range}(C^\intercal). \tag{4.22}$$

Then using $\text{null}(C) = \text{range}(G)$:

$$\mathbb{R}^E = \text{range}(G) \oplus \text{range}(C^\intercal). \tag{4.23}$$

Equation (4.23) means that there exists an r such that $Gr = f_t$, and there exists a v such that $C^\intercal v = f_c$. We have already proved r was unique. To prove that v is unique we use rank nullity. Equation eq. (4.23) guarantees $E = \text{rank}(G) + \text{rank}(C^\intercal)$. In general G has rank $m - 1$ since the Laplacian, $G^\intercal G$, has nullity equal to the number of connected components in the network [34]. We assumed the network is connected, so $G^\intercal G$ has nullity

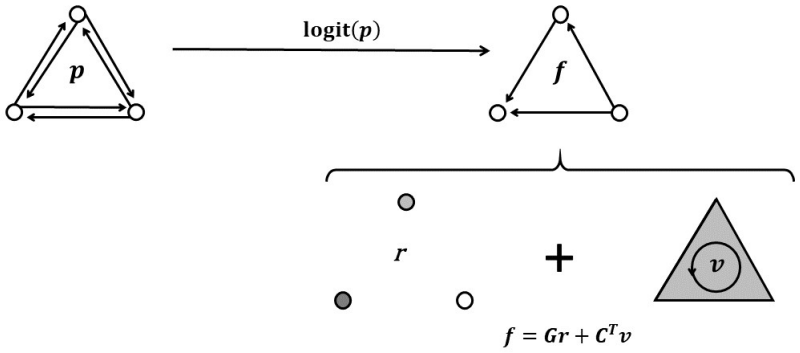

Figure 4.4: A schematic representation of the decomposition for a complete tournament on three competitors. The edge flow f is set equal to $\text{logit}(p)$, and then broken into a set of ratings r and vorticities v, such that $f = Gr + C^\intercal v$.

1, thus G has a one-dimensional nullspace. This nullspace corresponds to the vector of all ones, since the gradient of a constant is zero. Therefore $\text{rank}(C^\intercal) = E - (m-1) = L$. By construction, $C^\intercal$ has L columns, therefore $C^\intercal$ is full rank. It follows that the linear system $C^\intercal v = f$ has a unique solution if $f \in \text{range}(C^\intercal)$.[18] □

This proves that an arbitrary tournament can be decomposed into a perfectly transitive and a perfectly cyclic tournament, where the perfectly transitive tournament is specified by a set of ratings, and the perfectly cyclic tournament is specified by a set of vorticities. The ratings associated with the HHD are the Hodge ratings proposed by [16]. Figure 4.4 provides a schematic representing the decomposition.

The gradient G has exactly 2 nonzero entries per edge, so it becomes more sparse as the number of competitors increases. As a consequence, the decomposition can be performed efficiently, even for large, fully connected networks. Methods are discussed in [18, 16], and in Chapter 3.

[18] This result could also be obtained more intuitively as follows. Note that if C is defined with respect to a fundamental cycle basis, then by ordering the edges so that all of the chords are indexed before all of the edges in the tree, the operator C is a block matrix whose first $L \times L$ block is the identity. It follows that C is in row reduced echelon form and has rank L. The column rank of a matrix is its row rank so the rank of $C^\intercal$ is also L.

The intransitivity measure associated with the HHD is the size of the cyclic component $||f_c||_2$. Because the HHD is a decomposition onto orthogonal subspaces, this measure is equal to the distance from f to the closest perfectly transitive tournament. Therefore the Helmholtz-Hodge intransitivity measure is conceptually analogous to the Slater intransitivity measure [104], and its variants [100], [113], [114]. Similarly, the transitivity measure associated with the HHD is the size of the transitive component $||f_t||_2$, and is the distance from f to the closest perfectly cyclic tournament.

Note that these measures are continuous in p. This sets the measure associated with the HHD apart from classical methods which depend only on the direction of competition encoded in $\mathcal{G}_{\rightarrow}$ such as the Kendall [97] or Slater [104] measures. These methods are discrete in p. This distinction is important, since it means that the Helmholtz-Hodge measure distinguishes between the cases $p_{AB} = p_{BC} = p_{CA} = 0.99$ and $p_{AB} = p_{BC} = p_{CA} = 0.51$ (intransitivity 7.96 and 0.07 respectively). Using the discrete measures, these two tournaments are equally intransitive. Thus the Helmholtz-Hodge measure is distinguishes between strong and weak intransitive cycles, and so reflects the absolute strength of cyclic competition. The discrete measures reflect the relative strength of cyclic competition since they only depend on the sign of f, which depends on both f_c and f_t. If the transitive part is large then it may mask weaker cyclic competition when using a discrete measure. For example, if $p_{AB} = 0.99, p_{BC} = 0.99$ and $p_{CA} = 0.49$ then it is clear that the probability that C beats A is much larger than might be expected using any predictive rating of the competitors. However, in this example competition is transitive so all discrete measures of intransitivity would return their minimal value, 0. In contrast, the Helmholtz-Hodge measure returns intransitivity 5.29. These examples are illustrated in Figure 4.5 Normalizing the Helmholtz-Hodge measures by $||f||_2$ produces the equivalent relative measures: $||f_c||_2/||f||_2$ and $||f_t||_2/||f||_2$.

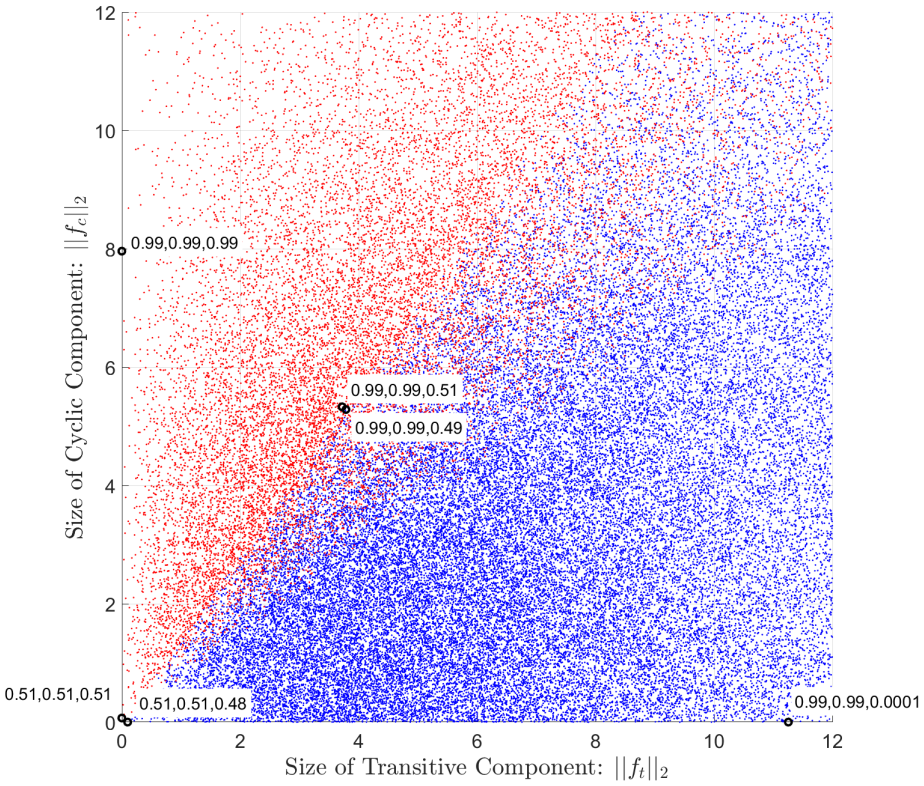

Figure 4.5: Transitivity and intransitivity of 10^4 triangular networks with randomly drawn win probabilities. The horizontal axis is the size of the transitive component and the vertical axis is the size of the cyclic component. Each scatter point is a sampled network. Blue scatter points are transitive, red are intransitive. The large black circles represent example networks. The text next to each example gives the probability A beats B, B beats C, and C beats A. If all of these numbers are greater than 0.5 then the network is intransitive. Note that the classification into transitive and intransitive draws a sharp distinction between networks whose win probabilities are nearly identical, while networks with similar win probabilities remain close to each other when using the Hodge measures. Also note that the boundary between transitive and intransitive networks is an angular sector, hence this classification is based on the relative sizes of the transitive and cyclic components, not their absolute sizes. In contrast, the Hodge measures reflect the absolute size of each component. Thus the example with win probabilities $0.99, 0.99, 0.49$ can be transitive and the example $0.51, 0.51, 0.51$ can be intransitive, even though the former has a larger cyclic component than the latter.

Comparison of the Hodge ratings and intransitivity measure to existing methods is provided in the Section 4.5.4.

Equivalent Formulations

Here we present six different approaches that arrive at the same decomposition. These provide different and useful perspectives on the HHD, and illustrate that it is robust to varying motivations. The ensuing Corollary follows directly from standard properties of projection onto orthogonal subspaces, so we omit the proof.

Corollary 19.1 (Equivalent Formulations)**.** *The following six decompositions are equivalent:*

1. $f = f_t + f_c$ *where* f_t *is arbitrage free and* f_c *is favorite free;*
2. $f = f_t + f_c$ *where* $f_t = Gr$ *for some rating* r *and* $f_c = C^\intercal v$ *for some vorticity* v*;*
3. *the ratings* r *satisfy:*

$$r = \text{argmin}_{u|\sum_i u_i=0} \left\{||Gu - f||_2^2\right\} \tag{4.24}$$

   and set $f_t = Gr$, $f_c = f - f_t$*;*

4. *the vorticities* v *satisfy:*

$$v = \text{argmin}_v\{||C^\intercal v - f||_2^2\} \tag{4.25}$$

   and set $f_c = C^\intercal v$, $f_t = f - f_c$*;*

5. *$f = f_t + f_c$ where $f_t = Gr$ for the unique ratings r such that the circulant $f - f_t$ is favorite free;*

6. *$f = f_t + f_c$ where $f_c = C^\intercal v$ for the unique vorticities v such that $f - f_c$ is arbitrage free.*

The first decomposition separates f into a pair of flows each defined by what it is not: namely, one is not circulatory, and the other has no tendency to diverge or converge. The second decomposition separates f into a pair of flows each defined by what they are: namely, one is perfectly transitive, and the other is perfectly cyclic. The equivalence of these two decompositions was established by theorem 19.

The next two decompositions are based on fitting problems. In each case the goal is to represent f as nearly as possible when restricted to the range of an operator. Decomposition 3 searches for a set of ratings r such the error, $Gr - f$, is minimized in the least squares sense. This means that the ratings produced by the HHD are a type of least squares rating, in particular, log least squares rating [96, 126, 127]. Least squares ratings methods are widely used [128, 129, 123, 130, 131, 132]. Comparisons are provided in the Section 4.5.4. Decomposition 3 also shows that the HHD is equivalent to finding the nearest perfectly transitive edge flow.

Similarly, Decomposition 4 searches for a set of vorticities v such that the error $C^\intercal v - f$ in approximating f with $C^\intercal v$ is minimized in the least squares sense. This is equivalent to finding the nearest perfectly cyclic edge flow. Although the literature has focused almost exclusively on Decomposition 3, decompositions 3 and 4 are dual to one another. This parity in approach sets the HHD apart from existing methods.

The final two decompositions are defined by enforcing a constraint on the residue when approximating f with either the gradient of a set of ratings or the curl transpose of a set of

vorticities. These approaches can be motivated as follows. Suppose one sought a rating r such that Gr approximated f. The error in this approximation (the circulant) is $Gr - f$. As long as the divergence of the circulant is nonzero the approximation has not captured a tendency of the edge flow to either point inwards towards, or outwards from, a competitor. If the net flow into a competitor is positive, then that competitor tends to outperform their neighbors in a way that the ratings fail to capture. Therefore it would be natural to adjust the ratings until the net flow into or out of any set of competitors is zero. That is, until the divergence of the circulant is zero, or equivalently, the circulant is favorite free.

The final decomposition can be motivated similarly. Define the *divergent*, $C^{\intercal}v - f$ to be the error upon approximating f with vorticity v. As long as the curl of the divergent is nonzero, the approximation has failed to capture some tendency of f to circulate. This tendency to circulate is exactly what the vorticities are meant to capture, so it is natural to look for a v such that the curl of the divergent is zero on every loop. That is, until the divergent is arbitrage free.

Corollary 19.1 shows that the decomposition into arbitrage free and favorite free components, perfectly transitive and perfectly cyclic components, the nearest perfectly transitive approximation, the nearest perfectly cyclic approximation, the perfectly transitive approximation with favorite free circulant/error, or the perfectly cyclic approximation with arbitrage free divergent/error, are all the same. The fact that the HHD is equivalent to all of these different approaches motivates its use.

4.5.3 Solution Methods

Corollary 19.1 casts the HHD as a pair of linear least squares problems:

$$
\begin{aligned}
f_t &= Gr, \quad r = \text{argmin}_{u \in \mathbb{R}^m | \sum_i u = 0} \{||Gu - f||_2^2\} \\
f_c &= C^\intercal v, \quad v = \text{argmin}_{w \in \mathbb{R}^L} \{||C^\intercal w - f||_2^2\}.
\end{aligned}
\tag{4.26}
$$

These least squares problems can be solved by standard numerical techniques. It can be helpful to recast the problem using the associated normal equations:

$$
\begin{aligned}
G^\intercal G r &= G^\intercal f \\
C C^\intercal v &= C f.
\end{aligned}
\tag{4.27}
$$

These are the discrete Poisson equations. They are analogous to the Poisson equations in continuous space, $\nabla^2 r(x) = \nabla \cdot f(x)$ and $\nabla^2 v(x) = \nabla \times f(x)$ which are used to perform the Helmholtz decomposition of a vector field $f(x)$ [16]. The matrix $G^\intercal G$ is the standard node Laplacian for the network. It is given by subtracting the unweighted $m \times m$ adjacency matrix A from the degree matrix D [19]:

$$
G^\intercal G = D - A. \tag{4.28}
$$

Therefore the ratings r associated with the HHD are solutions to the linear system, $(D - A)r = G^\intercal f$. Similar systems are used to solve for the Massey and Colley systems, only the Laplacian has to be adjusted to account for the number of observed events [128, 130, 123].

This equation is easy to interpret. Consider the i^{th} competitor. Suppose the i^{th} competitor has degree $d_i = |\mathcal{N}(i)|$. Let $\bar{r}_i = \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} r_j$ be the average rating of the neighbors of i. Then the left hand side of the discrete Poisson equation 4.27 is equal to the difference between the rating of competitor i, and the average rating of its neighbors, weighted by the number of neighbors: $[G^\intercal G r]_i = d_i(r_i - \bar{r}_i)$. We expect this to be positive when i is

[19] The adjacency matrix A has entries $a_{ij} = 1$ if i and j are connected and zero otherwise. The degree matrix D is diagonal, and has diagonal entries d_{jj} equal to the degree of competitor j, $|\mathcal{N}(j)|$.

better than its average neighbor, and negative when i is worse than its average neighbor. Note that this is equivalent to $[G^\intercal Gr]_i = \sum_{j \in \mathcal{N}(i)} (r_i - r_j)$ so the left hand side may also be interpreted as the net difference between the rating of i and the rating of all of i's neighbors. The i^{th} entry on the right hand side is the sum of all f_{ij} where $j \in \mathcal{N}_i$. This is positive if i tends to beat most of its neighbors, and negative if i tends to lose to most of its neighbors. Taking the second interpretation of the left hand side, and dividing both sides by d_i we see that the discrete Poisson equation requires that, for every i, the difference in rating between i and its average neighbor (the average difference in ratings) must equal the average flow into i.

This equation can also be solved efficiently. Note that the original equation $G^\intercal Gr = G^\intercal f$ does not have a unique solution since it does not enforce $\sum_i r_i = 0$. The easiest way to enforce this constraint is to arbitrarily set $r_1 = 0$. Then the corresponding row and column can be removed from the equation, which now has a unique solution. Then, after the equation is solved all of the ratings can be shifted so that the mean rating is zero. Once a row and column have been removed the fastest direct method is typically Cholesky decomposition since the node Laplacian is square and symmetric. If the tournament is sparse then iterative methods such as CGLS or LSQR converge quickly and are recommended by [16]. The iterative approach is also attractive since it only requires a routine for performing the product $[G^\intercal G]u$ for arbitrary $u \in \mathbb{R}^m$. This can be done efficiently since the action of the node Laplacian only requires knowledge of the adjacency structure (which competitor competes directly with which other competitors). This is the recommended method if the tournament is large and sparse.

If the network is complete then the discrete Poisson equation $G^\intercal Gr = G^\intercal f$ can be solved analytically. If there are m competitors competing in a complete tournament then:

$$r = \frac{1}{m} G^\intercal f. \tag{4.29}$$

Regardless the method used, once the Hodge ratings r are known then the perfectly transitive component $f_t = Gr$ can be easily computed. Similarly, once the perfectly transitive component is known the perfectly cyclic component f_c is simply given by $f - f_t$. Therefore, to find the two components it is enough to solve the discrete Poisson equation for the ratings[20]. This means that most of the decomposition can be performed without ever specifying a cycle basis.

To solve for the vorticities we may either solve the least squares problem 4.26, the associated discrete Poisson equation 4.27, or the overdetermined linear system $C^\intercal v = f_c$. The latter is guaranteed to have a solution since $f_c \in \text{range}(C^\intercal)$.

4.5.4 Comparison

Comparison of Hodge rating to Least Squares Rating Methods

When the tournament is arbitrage free then $Cf = 0$ so $f = Gr$ for Hodge ratings r. Theorem 15 established that r are consistent with the Elo rating and the Bradley-Terry ratings since the win probabilities $p_{ij} = \text{logistic}(r_i - r_j)$ and $p_{ij} = \frac{q_i}{q_i + q_j}$ where $q_i = e^{r_i}$. This close connection to the Bradley-Terry model is noted in [16].

When the tournament is not arbitrage free then there are no ratings r which satisfy $Gr = f$, so the Hodge ratings are the solution to the unweighted least squares problem 4.27. Since the Hodge ratings r minimize the least squares error between Gr and f the Hodge ratings are equivalent to the log least squares approach used by [96, 126, 127, 141].

[20] Note that the same is true of the vorticities, but there are more applications in which the ratings are of primary interest. Moreover the Poisson equation for the ratings is $m - 1$ dimensional while the Poisson equation for the vorticities is L dimensional, and in most applications $m < L$.

As noted before this is equivalent to using an unweighted Massey rating system with $f_{ij} = \text{logit}(p_{ij})$. Weighting by the number of games produces a Massey type rating system, while weighting by the number of games and regularizing to account for the Laplace rule of succession produces the Colley rating system [128, 123, 130]. It can be shown that these two systems can be derived from an appropriate choice of prior when the win probabilities are estimated from an observed series of wins and losses.

When the tournament is complete the Hodge ratings r are given by $r = \frac{1}{m} G^{\intercal} f$, or, examining one term at a time:

$$r_i = \frac{1}{m} \sum_{j \neq i}^{m} f_{ij} = \frac{1}{m} \sum_{j} \log\left(\frac{p_{ij}}{p_{ji}}\right) = \log\left(\left(\frac{\prod_{j \neq i}^{m} p_{ij}}{\prod_{j \neq i}^{m} p_{ji}}\right)^{1/m}\right). \tag{4.30}$$

Therefore, when the tournament is complete the rating of the i^{th} competitor is equal to their average log-odds against a randomly drawn opponent, including themselves [21]. This is, equivalently, the log of the geometric average of their odds against a randomly drawn opponent. Therefore $q_i = e^{r_i}$ may be interpreted as the geometric average of the odds that competitor i beats a uniformly drawn opponent [126]. Alternatively, this may also be interpreted as the log of the odds that i beats all other competitors, divided by m.

Note that the odds that i beats all other competitors depends on the ratio of the probability that i beats all of its neighbors to the probability that i loses to all of its neighbors. These are the two probabilities that appeared in the neighborhood condition that defines favorite free tournaments (see Equation (4.10)) . If a tournament is favorite free then these two probabilities are equal, so the log of the ratio is zero. It follows that if a tournament is complete and favorite free then the Hodge ratings are all zero.

If the tournament is not complete then it is not true that $r = \frac{1}{m} G^{\intercal} f$, however it is still

[21] The odds that they beat themselves are one to one so the log-odds are zero

true that r satisfy the discrete Poisson equation $G^\intercal G r = G^\intercal f$. Therefore:

$$r_i - \bar{r}_i = \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} f_{ij} = \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} \log\left(\frac{p_{ij}}{p_{ji}}\right) = \log\left(\left(\frac{\prod_{j \in \mathcal{N}(i)}^{m} p_{ij}}{\prod_{j \in \mathcal{N}(i)}^{m} p_{ji}}\right)^{1/d_i}\right). \tag{4.31}$$

This means that, when a tournament is not complete, the difference in the rating of the i^{th} competitor, and the average rating of its' neighbors, is equal to the average log-odds that competitor i beats a uniformly drawn opponent from its neighborhood. This means that the only difference in interpretation when moving from a complete tournament to an arbitrary tournament is that the ratings of a competitor i in an arbitrary tournament are given by the average log-odds that they beat a randomly drawn opponent from their neighborhood, plus the average rating of their neighbors. The addition of the average rating of their neighbors accounts for the strength of their neighborhood/schedule.

As in a complete tournament this can be reinterpreted as the average rating of the neighborhood plus the log of the geometric average of the odds that i beats a randomly drawn neighbor, or, plus the log of the odds that i beats all neighbors scaled by the size of the neighborhood. Also as before, if the tournament is favorite free then the log-odds that i beats all of its neighbors is equal to zero, so in a favorite free tournament the rating of every competitor is equal to the average rating of their neighbors. Therefore, if a tournament is favorite free then the Hodge rating of every competitor is equal to zero.

Comparison of Hodge intransitivity to existing Intransitivity Measures

The Helmholtz-Hodge intransitivity measure is defined to be the distance between the true edge flow f and the nearest perfectly transitive edge flow, where distance is measured by the l_2 metric. Lemma 6 corollary 19.1 guaranteed that this is equivalent to the two-norm of

the perfectly cyclic component of f.

Clearly this measure is analogous to the Slater measure defined in Equation (4.6), which was the distance between a competitive network and the nearest transitive network, with distance measured in the number of competitive reversals [104]. Like the Slater measure, the Helmholtz-Hodge measure is associated with a nearest transitive tournament. When considering only the sign of $p_{ij} - 1/2$ the nearest transitive tournament was the Kemeny optimal tournament, and was associated with a Kemeny optimal ranking. Since the nearest perfectly transitive tournament to the original tournament is given by the perfectly transitive component of the HHD, the Hodge ratings r may be thought of as analogous to the optimal ratings, since they produce the perfectly transitive edge flow closest to the true edge flow. Therefore the Helmholtz-Hodge intransitivity measure can be reasonably considered a member of the family of "distance to nearest transitive" intransitivity measures which included the Slater [104], Petraitis [100], and Ulrich measures [114].

Unlike the Slater measure the Hodge-Helmoltz intransitivity measure is:

1. associated with a unique nearest perfectly transitive tournament, specified by the Hodge ratings,
2. can be applied to any finite, connected, reversible tournament,
3. is efficiently computable for large networks,
4. is analytically computable for complete tournaments and is analogous to the Landau measure in this case,
5. and is continuous in p and f.

The first point is guaranteed by the uniqueness of the decomposition. The following three points are all established by the fact that the measure can be computed directly from

the ratings, $\text{Int}_{\text{H}}(f) = ||f_c||_2 = ||Gr - f||_2$, and the ratings can be solved for efficiently by solving the discrete Poisson equations, or the original least squares problem.

The numerical methods for solving for the ratings scale well, and are very efficient when the tournament is sparse. Since the HHD has a unique solution for all finite connected reversible tournaments, and since the same numerical methods can be applied in all cases, the Helmholtz-Hodge intransitivity measure generalizes to all finite connected reversible tournaments.

If the tournament is complete then the ratings are given by $r = \frac{1}{m}G^{\intercal}f$ so $f_c = (I - \frac{1}{m}GG^{\intercal})f$ and:

$$\text{Int}_{\text{H}}(f) = ||(I - \frac{1}{m}GG^{\intercal})f||_2 \tag{4.32}$$

In the complete case the intransitivity measure is analogous to the Landau measure since $||f_t||_2^2 = f_t^{\intercal}f_t = (f_t + f_c)^{\intercal}f_t = f^{\intercal}f_t = \frac{1}{m}f^{\intercal}GG^{\intercal}f = \frac{1}{m}\sum_{i=1}^{m}[G^{\intercal}f]_i^2$ [22]. This is the variance in the divergence of f evaluated at each node since the mean value of the divergence is zero. The mean value of the divergence is zero since $\frac{1}{m}\sum_{i=1}^{m}[G^{\intercal}f]_i = \frac{1}{m}1^{\intercal}G^{\intercal}f$ where 1 is the vector of all ones. But $1^{\intercal}G^{\intercal} = G1$ and the gradient of a constant is zero. Therefore $||f_t||_2^2$ is the variance in the divergence of the edge flow. For a given f the largest the variance in the divergence of f could be is $||f||_2^2$ (when f is perfectly transitive). Therefore, if we normalize the transitivity by its largest possible value then this normalized transitivity is the variance in the divergence, $||f_t||_2^2$, divided by the largest this variance could possibly be, $||f||_2^2$. The corresponding intransitivity is one minus this ratio. In comparison the Landau measure is one minus the ratio of the variance in the in-degree of each node to the largest that this in-degree could be. The in-degree of each node is analogous to the divergence since, if the edge flow is rounded to $1/2$ when $f_k > 0$ and $-1/2$ when $f_k < 0$, then the

[22] since $f_c^{\intercal}f_t = vCGr$ and $CG = 0$.

variance in the in-degree is the variance in the rounded edge flow.

The final point, that the Helmholtz-Hodge measure is continuous in p, is obvious from the fact that f is a continuous function of p, the perfectly cyclic component f_c is an orthogonal projection of f onto a fixed subspace, and the two norm of f_c is continuous in the entries of f_c. This is an important point since it means that the Helmholtz-Hodge measure distinguishes between $p_{AB} = p_{BC} = p_{CA} = 0.99$ and $p_{AB} = p_{BC} = p_{CA} = 0.51$. Using the discrete measures these two tournaments are equally intransitive. Using the Hodge-Helmoholtz measure the former has intransitivity 7.96 and the latter has intransitivity 0.07. This means that the Helmholtz-Hodge measure is capable of distinguishing between strong and weak intransitive cycles, so reflects the absolute strength of cyclic competition. The discrete measures reflect the relative strength of cyclic competition since they only depend on the sign of f, which depends on both f_c and f_t. If f_t is very small relative to f_c on every edge then the discrete measures will all return their maximum possible values, indicating strong intransitivity, even if f_c is itself very small (as in the example given above). Alternatively, if the transitive part is large then it may mask weaker cyclic competition when using a discrete measure. For example, if $p_{AB} = 0.99, p_{BC} = 0.99$ and $p_{CA} = 0.49$ then competition is transitive so all discrete measures of intransitivity would return their minimal value, 0. However it is clear from the probabilities that the probability that C beats A is much, much larger than might be expected by any reasonable rating of the competitors given that A beats B 99 out of 100 events and B beats C 99 out of 100 events. In contrast, the Helmholtz-Hodge measure returns intransitivity 5.29. Note that this is larger than the value returned for the intransitive loop with win probabilities 0.51.

Therefore the Helmholtz-Hodge measure may be viewed as a measure of the *absolute* strength of cyclic competition, as it measures the size of the perfectly cyclic component independent of the size of the transitive component. This means that it is not necessarily

zero even if the network is transitive. This should be no surprise as not all transitive tournaments are perfectly transitive, and the Helmholtz-Hodge measure is the distance between the given tournament and the nearest perfectly transitive tournament. In effect, the intransitivity measure associated with the HHD measures how closely the ratings r can predict the probabilities via $p_{ij} = \text{logistic}(r_i - r_j)$.

4.6 The Trait-Performance Theorem

How intransitive is a typical tournament? Using the intransitivity measure associated with the HHD, this question is the same as asking, how cyclic is a tournament on average?

Answering this question requires defining a statistical model for sampling tournaments - in particular, for sampling edge flows. How do assumptions about the distribution of possible edge flows affect the expected strength of cyclic competition? What statistical features tend to promote or suppress cyclic competition?

We initially explore these questions for a generic null model in which the edge flow, F, is sampled randomly from an unspecified distribution. This analysis identifies which statistical features of the edge flow, and which features of the network topology, influence the expected strength of cyclic competition. This sets the stage for our main result. If the edge flow is sampled using a trait-performance model, then the correlation structure of the edge flow takes on a canonical form which depends only on *two* statistical quantities: the variance in the flow on each edge, and the correlation in the flow on pairs of edges that share an endpoint. This simplified correlation structure allows us to express the expected sizes of the cyclic and transitive components in a simple closed form that separates the influence of the network topology from the chosen trait-performance model.

4.6.1 Generic Null Models

We start by considering a generic null-model for the edge flows f. Let $F \in \mathbb{R}^E$ be a random edge flow drawn from some distribution. Assume that the expected edge flow $\bar{f} = \mathbb{E}[F]$ is known, as is the covariance $V = \mathbb{E}[(F - \bar{f})(F - \bar{f})^\intercal]$.

Let P_c be the orthogonal projector onto the space of perfectly cyclic (favorite free) tournaments. Then the expected absolute strength of cyclic competition is:

$$\begin{aligned} \mathbb{E}[||F_c||^2] = \mathbb{E}[F^\intercal P_c^\intercal P_c F] = \mathbb{E}[F^\intercal P_c F] = \mathbb{E}\left[\sum_{kl} (P_c)_{kl} F_k F_l\right] = \\ \sum_{kl} (P_c)_{kl} \mathbb{E}[F_k F_l] = \sum_{kl} (P_c)_{kl} (\bar{f}_k \bar{f}_l + v_{kl}) = ||\bar{f}_c||^2 + \text{trace}(P_c V) \end{aligned} \tag{4.33}$$

where $||\bar{f}_c||^2 = \bar{f}^\intercal P_c \bar{f}$ and $\text{trace}(P_c \Sigma) = \sum_{kl} (P_c)_{kl} v_{kl}$ is the matrix inner product between the projector and the covariance matrix.

Therefore, no matter the underlying distribution of edge flows, the expected strength of cyclic competition is determined exclusively by three quantities: the *expected edge flow*, the *covariance in the edge flow*, and the *topology of the network* (which determines P_c).

The matrix inner product can be simplified if the flows on each edge are independent. Then V is diagonal with entries $\sigma_k^2 = \mathbb{E}[(F_k - \bar{f}_k)^2]$. It follows that $\text{trace}(P_c V) = \sum_{k=1}^{E} (P_c)_{kk} \sigma_k^2$.

The nonzero eigenvalues of a projector all equal one, so its trace equals the dimension of the space it projects onto. The projector P_c projects onto the space of perfectly cyclic tournaments, which has dimension $L = E - (m - 1)$. Therefore $\sum_k (P_c)_{kk} = L$. Rewrite the expected strength of cyclic competition:

$$\mathbb{E}[||F_c||^2] = ||\bar{f}_c||^2 + L \sum_{k=1}^{E} \left(\frac{(P_c)_{kk}}{L}\right) \sigma_k^2. \tag{4.34}$$

Since the diagonal entries of an orthogonal projector are always nonnegative, the right hand term can be interpreted as a weighted average of the variance on each edge. Therefore, when the edges are independent, the expected strength of cyclic competition is given by the strength of the cyclic component of the expected edge flow, plus the dimension of the loop space times a weighted average of the variance on each edge. Similarly, the expected strength of transitive competition is:

$$\mathbb{E}[||F_t||^2] = ||\bar{f}_t||^2 + (m-1)\sum_{k=1}^{E}\left(\frac{(P_t)_{kk}}{m-1}\right)\sigma_k^2 \tag{4.35}$$

and the expected total strength of competition is:

$$\mathbb{E}[||F||^2] = ||\bar{f}||^2 + E\bar{\sigma}^2 \tag{4.36}$$

where $\bar{\sigma}^2$ is the average of the variance in the flow on each edge. Equation eq. (4.36) is valid even if the edges are not independent, as the projector onto the full space is simply the identity.

Equations eq. (4.34) - eq. (4.36) show that the contribution to the expected strength of competition from the variances is not distributed equally between the transitive and cyclic spaces. Instead, the amount that is cyclic is proportional to the number of cycles, while the amount that is transitive is proportional to the number of competitors. As a result, adding edges to a network will typically increase the expected degree to which competition is cyclic. It follows that sparse networks with randomly drawn edge flows will be relatively more transitive than would be expected given $\bar{f}$, while dense networks will typically be more cyclic. It also follows that, for a posterior distribution of possible edge flows given observed data, uncertainty will likely lead to an overestimate of the degree to

which competition is cyclic, if the network is dense.

Further simplifications emerge when a network is edge-transitive or has homogeneous variances σ_k^2. A network is edge-transitive if the edges are indistinguishable once the node labels are removed. This symmetry implies that p_{kk} is independent of k, regardless the space the projector maps onto. Therefore $(P_c)_{kk} = L/E$ and $(P_t)_{kk} = (m-1)/E$. Thus:

$$\begin{aligned}
\mathbb{E}[||F_c||^2] &= ||\bar{f}_c||^2 + L\bar{\sigma}^2 \\
\mathbb{E}[||F_t||^2] &= ||\bar{f}_t||^2 + (m-1)\bar{\sigma}^2 \\
\mathbb{E}[||F||^2] &= ||\bar{f}||^2 + E\bar{\sigma}^2
\end{aligned} \tag{4.37}$$

where $\bar{\sigma}^2 = \frac{1}{E}\sum_{k=1}^{E} \sigma_k^2$.

Any symmetric network, or complete network, is edge-transitive, so these equations apply to all symmetric networks and all complete networks with edge flows drawn independently on each edge. Alternatively, if the variances σ_k^2 do not depend on k, then any weighted average of the variances is equal to $\bar{\sigma}^2$. In this case equations eq. (4.37) also apply.

These results show that, in general, the expected strengths of cyclic and transitive competition depend on the expected edge flow, the uncertainty in the edge flow, and the topology of the network. Increasing the uncertainty in the edge flow increases the expected strength of both cyclic and transitive competition, but does not increase both equally. If the graph is sparse, then increasing the uncertainty will typically promote transitive competition more than cyclic. If the graph is dense, then increasing the uncertainty will typically promote cyclic competition more than transitive. If a tournament is complete, then $E = m(m-1)/2$ so $(m-1)/E = 2/m$ and $L/E = 1 - 2/m$. It follows that for a complete tournament with more than four competitors, any uncertainty in the edge flow

will typically bias competition to appear more cyclic than transitive. This is necessarily true if the edges are drawn independently, and the graph is either edge-transitive or the variances on each edge are all the same.

Numerical studies have suggested that filling in missing edges with randomly drawn F typically overestimates the degree to which competition is cyclic [95]. Our result provides a rigorous explanation for this observation. When the edge flow F is drawn randomly to fill in missing data, it is usually drawn independently and identically distributed, cf. [143]. From equation eq. (4.37) it is clear that if edges are added until the network is complete, then, for any tournament with more than four competitors, the resulting “imputed" tournament will likely be significantly more cyclic than the original tournament. Therefore, unless the edge flows are well-modeled by assuming that the F_k are independent and identically distributed, *and* that all pairs of competitors could compete, this procedure is not valid for estimating the strength of cyclic competition in a partially observed tournament.

The simplified equations eq. (4.34) - eq. (4.37) are valid only if the edge flows are drawn independently, which is rarely the case for real-world tournaments. When the edge flows are not drawn independently, the edge flow covariance matrix is not diagonal, and the simplification leading from Equation (4.33) to Equation (4.34) no longer holds. This makes it more challenging to identify how the topology of the network promotes or suppresses cyclic competition. Nevertheless, as we show in the next section, using a more principled model for sampling F, ensures that the covariance matrix V takes on a canonical form. This form clarifies the interaction between the topology of the network and the distribution of edge flows.

4.6.2 Trait-Performance

The outcomes of real-world competition events are typically influenced by a constellation of underlying traits of the competitors. Examples of trait-based competition models abound, ranging from sports[23] to biology. [24] In some cases, trade-offs inherent in certain traits have been observed to lead to cyclic competition between organisms [102, 103].[25] In such examples, trade-offs lead to advantages against certain opponents, and weaknesses that are exploited by others. In evolutionary biology, trade-offs of this kind challenge the notion that members of intransitive communities can be consistently ranked according to fitness. Intransitivity can lead to deeply counterintuitive evolutionary dynamics [151, 152], and may promote biodiversity since no single species has an absolute advantage over all competitors [101, 105, 106, 107, 113]. These considerations motivate a study of how demographics (the distribution of traits), and the way traits confer success, either promote or suppress cyclic competition.

Therefore, we now suppose that win probabilities p can be modeled as a function of

[23] Some predictive tennis models estimate the probability that one competitor will beat another based on a parameterized model for the probability that each player will win a point, where the underlying parameters depend on traits of the players [149]. Similarly, considerable effort has been devoted to predictive models for baseball based on team and player statistics [121].

[24] Ecological studies of competition for dominance in social hierarchies have analyzed how traits confer success, because selection acts on heritable traits contributing to reproductive success. Examples include competition among male northern elephant seals [119] and male Cape dwarf chameleons [118]. Relevant traits for elephant seals include body mass, length, age, and time of arrival on the beach [119]. Relevant traits for chameleons include body mass, length from snout to base of tail, length of the tail, jaw length, head width, casque size, and size of a pink colored flank patch used in signaling [118].

[25] Two particularly famous examples are side-blotched lizards and colicin producing *E. coli* [102, 103]. In the former example, large orange-throated males maintain large territories, medium blue-throated males defend small territories, while small yellow-throated 'sneaker' males resemble females and do not maintain territories. Orange-throated males typically defeat the smaller blue-throated males, who defeat the even smaller yellow throated males, who defeat the orange throated males by sneaking into their territories [103]. In the latter example, three strains of *E. coli* were grown in direct competition in a laboratory setting. The first strain produced a colicin toxin, the second was susceptible to the toxin, and the third was resistant to the toxin but not toxin-producing. In the absence of the resistant strain, the toxic strain could outcompete the susceptible strain. In the absence of the toxic strain, the susceptible strain could outcompete the resistant strain, which reproduced more slowly because resistance is costly. But, in the absence of the susceptible strain, the resistant strain could outcompete the toxic strain by reproducing more quickly [102].

some underlying traits x of each competitor. Let $X(i) = [X_1(i), \ldots, X_T(i)]$ denote the T randomly sampled traits of the i^{th} competitor. Then let $f(x, y)$ be a performance function, such that $f(x, y)$ is the log-odds that a competitor with traits x would beat a competitor with traits y.

To construct a trait-performance model assume that:

1. The trait vectors of the competitors are drawn independently and identically from a trait distribution π_x.

2. There exists a performance function $f(x, y)$ that maps from $\mathbb{R}^\intercal \times \mathbb{R}^\intercal$ to $\mathbb{R}$. We require that the performance function is alternating $f(x, y) = -f(y, x)$ for any trait vectors x and y in the support of π_x. This ensures that f can be used to generate an edge flow. It also ensures that the performance function is fair, $\mathbb{E}[f(X, Y)] = 0$, since if X and Y are drawn i.i.d then $\mathbb{E}[f(X, Y)] = \mathbb{E}[f(Y, X)] = -\mathbb{E}[f(X, Y)]$ which implies $\mathbb{E}[f(X, Y)] = 0$.

3. There exists a connected competitive network $\mathcal{G}_{\rightleftarrows}$ with edges representing possible competition events, and the network is either fixed a priori or sampled independently from the traits.

Assumptions 1 and 3 are the most restrictive. The first assumes all competitors are drawn from the same demographic pool. Different demographic pools can be incorporated into the model by adding a trait which indexes which pool each competitor is sampled from, provided that trait can be sampled independently of the graph. For example, Major League Baseball team budgets vary widely. In 2018 the Yankees' total value was over 4.6 billion dollars, which was more than the total value of the bottom six teams combined [153]. This difference resoucres gives high value teams the opportunity to pay higher salaries[26] and

[26]For example, in 2019 the Yankees' combined payroll was three times larger than the Marlins'.

thus attract star players. Thus rich teams are in a different demographic pool than poor teams, so the wealth of the teams could be incorporated as one of their traits.

The third assumption treats the network topology (who competes with whom) independently from the traits of the competitors. This may not be realistic if competitors avoid competing when they are likely to lose [142]. This also limits our ability to model systems where traits are heritable, or distributed differently across different clusters of competitors (different divisions, or local populations).

The second assumption is the least restrictive since it is valid whenever the probability that one competitor beats another can be conditioned on the traits of the competitors, independent of their location on the network.

Under these assumptions, we define a trait-performance model as follows. First, sample $X(i) \sim \pi_x$ for all competitors i. Then, set $F_k = f(X(i(k)), X(j(k)))$, where $i(k), j(k)$ are the endpoints of edge k.

Theorem 20 (Trait-Performance)**.** *Let $\mathcal{G}_{\rightleftarrows}$ be a competitive network satisfying assumption 3. If the traits of each competitor are drawn independently from π_x, and the edge flow is defined by $F_k = f(X(i(k)), X(j(k)))$ where $f(x, y)$ is an alternating performance function, then the covariance V of the edge flow has the form:*

$$V = \sigma^2 \left[I + \rho \left(G G^{\intercal} - 2I \right) \right] \tag{4.38}$$

where σ^2 is the variance in F_k for arbitrary k, and ρ is the correlation coefficient between $f(X, Y)$ and $f(X, W)$ for X, Y, W drawn i.i.d from π_x.

Moreover:

$$\mathbb{E}\left[\frac{1}{E}||F||^2\right] = \sigma^2 \xrightarrow{\textit{decompose}} \begin{cases} \mathbb{E}\left[\frac{1}{E}||F_t||^2\right] = \sigma^2\left[\frac{(m-1)}{E} + 2\rho\frac{L}{E}\right] \\ \mathbb{E}\left[\frac{1}{E}||F_c||^2\right] = \sigma^2\left(1-2\rho\right)\frac{L}{E} \end{cases} \tag{4.39}$$

Therefore, the expected absolute strength of competition is independent of ρ, the size of the transitive component is monotonically increasing in ρ, and the size of the cyclic component is monotonically decreasing in ρ. The correlation ρ ranges from 0 to $1/2$, and if $\rho = 1/2$ then competition is perfectly transitive.

Proof. First consider the covariance matrix V.

Since the trait vectors are drawn i.i.d from the trait distribution, the diagonal entries of the covariance are given by:

$$V_{kk} = \mathbb{E}\left[(f(X(i(k), X(j(k)))^2\right] = \mathbb{E}\left[(f(X,Y))^2\right] \equiv \sigma^2 \tag{4.40}$$

where X, Y are drawn i.i.d from the trait distribution, and σ^2 is the variance in $f(X,Y)$. Thus, the diagonal entries of the covariance are identical.

The off-diagonal entries are $V_{kl} = \mathbb{E}\left[f(X(i(k)), X(j(k))) \cdot f(X(i(l)), X(j(l)))\right]$.

Suppose the edges k and l do not share an endpoint. Then $i(k) \neq i(l)$ or $j(l)$ and $j(k) \neq i(l)$ or $j(l)$. Then $f(X(i(k)), X(j(k)))$ is a function of two random vectors, and $f(X(i(l)), X(j(l)))$ is a function of two other random vectors, where the pair of random vectors are independent. It follows that $f(X(i(k)), X(j(k)))$ is independent of $f(X(i(k)), X(j(k)))$. Then, since competition is fair for all alternating performance functions $V_{kl} = \mathbb{E}\left[f(X(i(k)), X(j(k))) \cdot f(X(i(l)), X(j(l)))\right]$ which, by independence, equals $\mathbb{E}\left[f(X(i(k)), X(j(k)))\right]\mathbb{E}\left[f(X(i(l)), X(j(l)))\right] = 0$. It follows that the support of the

covariance matches the adjacency structure of the edges of the competition network.

If the edges do share an endpoint, then there are four possibilities. Either $i(k) = i(l)$, $j(k) = j(l)$, $i(k) = j(l)$, or $j(k) = i(l)$. We say that the edges are *consistently oriented* if they share either the same starting point or the same ending point, and are *inconsistently oriented* if the endpoint of one is the start of another. Since all the trait vectors are drawn i.i.d., we suppress the indices and let the three trait vectors Y, W, Z be drawn i.i.d. from π_x. The performance function is alternating, so:

$$\begin{aligned} \mathbb{E}[f(Y,W)f(Y,Z)] &= \mathbb{E}[f(W,Y)f(Z,Y)] \equiv \rho\sigma^2 \\ \mathbb{E}[f(Y,W)f(Z,Y)] &= \mathbb{E}[f(W,Y)f(Y,Z)] = -\mathbb{E}[f(Y,W)f(Y,Z)] = -\rho\sigma^2 \end{aligned} \tag{4.41}$$

where ρ is the correlation coefficient between $f(Y,W)$ and $f(Y,Z)$. Notice that a positive correlation indicates that the probability that A beats B is increased by conditioning on the event that A beats C.

The edge graph is the graph with a node for each edge in the competition network, and with an undirected edge between nodes corresponding to connected edges in the competition network (Figure 4.6). Let A_E be the weighted adjacency matrix for the edge graph with $a_{Ekl} = +1$ or -1 if edges k and l are consistently or inconsistently oriented with respect to a shared endpoint. Then:

$$V = \sigma^2 \left[I + \rho A_E \right]. \tag{4.42}$$

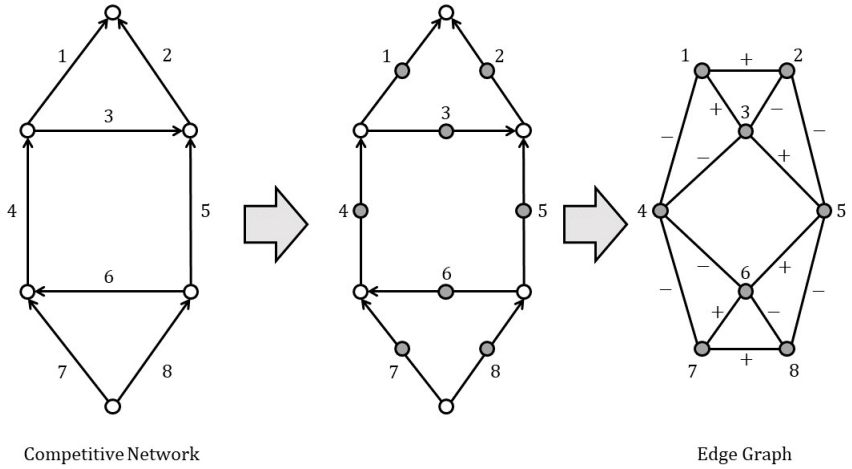

Figure 4.6: The edge graph (right) associated with a competitive network (left). The middle panel shows an intermediate graph where a node has been introduced for each edge. The edges of the competitive network become the nodes of the edge graph. The edges of the edge graph correspond to nodes in the competitive network that are the shared endpoint of a pair of edges. These are labeled with a $+$ or $-$ to indicate whether the edges are consistently or inconsistently oriented with respect to the shared endpoint.

The weighted adjacency matrix A_E for the edge graph is equal to $GG^\intercal - 2I$ since:

$$[GG^\intercal]_{kl} = (e_{i(k)} - e_{j(k)})^\intercal (e_{i(l)} - e_{j(l)}) = \left\{ \begin{array}{l} 2 \text{ if } k = l \\ 1 \text{ if } i(k) = i(l) \text{ or } j(k) = j(l) \\ -1 \text{ if } i(k) = j(l) \text{ or } j(k) = i(l) \\ 0 \text{ else} \end{array} \right\} \tag{4.43}$$

where $e_i \in \mathbb{R}^m$ is the indicator vector for node i. Thus we establish eq. (4.38).

All of the absolute measures of the strength of competition (squared) are given by the squared length of the orthogonal projection of the edge flow onto some subspace. Let P_S be an arbitrary orthogonal projector onto some subspace S. By construction, the edge flow is zero mean, therefore, by equation eq. (4.33), the expected value of the associated measure

is:

$$\mathbb{E}\left[||F_S||^2\right] = \text{trace}(P_S V) \tag{4.44}$$

where V is the covariance matrix of the edge flow F.

The intensity of competition, $||F||^2$, corresponds to the projector I, $||F_t||^2$ corresponds to the projector P_t, and $||F_c||^2$ corresponds to the projector P_c. Then, by equation eq. (4.44):

$$\mathbb{E}\left[\frac{1}{E}||F||^2\right] = \frac{1}{E}\text{trace}(V) = \frac{E}{E}\sigma^2 = \sigma^2. \tag{4.45}$$

This formula establishes that the absolute strength of competition only depends on the variance σ^2 in each individual performance function.

To compute $||F_t||^2$, use equation eq. (4.44) with projector P_t:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{E}||F_t||^2\right] &= \frac{1}{E}\text{trace}(P_t V) = \frac{\sigma^2}{E}\text{trace}\left(P_t[I + \rho(GG^\intercal - 2I)]\right) \\ &= \frac{\sigma^2}{E}\text{trace}\left(P_t\right) + \frac{\rho\sigma^2}{E}\text{trace}\left(P_t(GG^\intercal)\right) - \frac{2\rho\sigma^2}{E}\text{trace}\left(P_t\right).\end{aligned} \tag{4.46}$$

The trace of an orthogonal projector equals the dimension of the subspace it projects onto, so $\text{trace}(P_t) = m - 1$. The range of $GG^\intercal$ is in the range of G, which is the subspace P_t projects onto. It follows that $P_t GG^\intercal = GG^\intercal$ so $\text{trace}(P_t GG^\intercal) = \text{trace}(GG^\intercal) = 2E$ (see equation eq. (4.43)). Therefore:

$$\mathbb{E}\left[\frac{1}{E}||F_t||^2\right] = \sigma^2\left[\frac{m-1}{E} + 2\rho\frac{E-(m-1)}{E}\right] = \sigma^2\left[\frac{m-1}{E} + 2\rho\frac{L}{E}\right]. \tag{4.47}$$

Since $L \geq 0$, $\mathbb{E}[\frac{1}{E}||F_t||^2]$ increases monotonically in ρ: the larger ρ, the more A beating B is correlated with A beating C, implying transitive competition.

To compute the expected absolute strength of cyclic competition (squared) we take

advantage of the orthogonality of the decomposition $f = f_c + f_t$:

$$\mathbb{E}\left[\frac{1}{E}||F_c||^2\right] = \mathbb{E}\left[\frac{1}{E}||F||^2\right] - \mathbb{E}\left[\frac{1}{E}||F_t||^2\right] = \sigma^2\left[1-2\rho\right]\frac{L}{E}. \tag{4.48}$$

It follows that the expected absolute strength of cyclic competition is monotonically decreasing in the correlation coefficient ρ. Note that, as when considering the generic null models, dense networks promote cyclic competition.

To conclude we show that $\rho \in [0, 1/2]$, so the expected measures are maximized and minimized when ρ is 0 or 1/2, respectively.

The correlation ρ is nonnegative since W and Z are i.i.d., thus $f(y, W)$ and $f(y, Z)$ are also i.i.d., so:

$$\begin{aligned}\sigma^2\rho &= \mathbb{E}_{Y,W,Z}[f(Y,W)f(Y,Z)] = \int_{\mathbb{R}^\intercal}\mathbb{E}_{W,Z}[f(y,W)f(y,Z)]\pi_x(y)dy \\ &= \int_{\mathbb{R}^\intercal}\mathbb{E}_W[f(y,W)]\mathbb{E}_Z[f(y,Z)]\pi_x(y)dy = \int_{\mathbb{R}^\intercal}\mathbb{E}_W[f(y,W)]^2\pi_x(y)dy \geq 0\end{aligned} \tag{4.49}$$

Here expectation is taken with respect to the variables in the subscript.

To prove that $\rho \leq 1/2$, note that all covariance matrices are positive semi-definite, so, for any vector u:

$$u^\intercal V u = \sigma^2 u^\intercal(I + \rho(GG^\intercal - 2I))u = \sigma^2(1-2\rho)||u||^2 + \rho u^\intercal GG^\intercal u \geq 0. \tag{4.50}$$

If $E > m-1$, then the network has at least one loop, so the range of $C^\intercal$ is non-empty, hence the nullspace of $G^\intercal$ is non-empty. Choosing u perfectly cyclic sets $G^\intercal u = 0$ so $\sigma^2(1-2\rho)||u||^2 \geq 0$ which requires $\rho \leq \frac{1}{2}$. If $E = m-1$ then the network is a tree, so all competition is necessarily perfectly transitive.

It follows that the expected absolute strength of *transitive* competition is minimized

when $\rho = 0$, and maximized when $\rho = 1/2$. In contrast, the expected strength of *cyclic* competition is maximized when $\rho = 0$, and minimized when $\rho = 1/2$.

If $\rho = 1/2$ then $\mathbb{E}[||F_c||^2] = 0$. The measure is nonnegative for all edge flows. Therefore, its expected value is only zero if the probability that $||F_c||^2 \neq 0$ is zero. In this case, the tournament is arbitrage free. It follows that, if $\rho = 1/2$, then the tournament must be perfectly transitive.[27] □

Theorem 20 establishes that the expected degree to which competition is transitive or cyclic depends principally on the density of the network, and the correlation structure of F. In particular, the degree to which a network is cyclic or transitive depends on the correlation between the performance of A against B with the performance of A against C. The larger this correlation, the more consistently each competitor performs, hence the more consistent the network is with a set of ratings.

The variance σ^2 and the correlation coefficient ρ could be computed given an assumed trait distribution π_x and performance function $f(x, y)$. This could be done analytically if π_x and f lead to simple calculations. Otherwise, σ^2 and ρ can be approximated numerically. The analytic method follows.

Suppose that X, Y are drawn from a sample space Ω which is a subset of $\mathbb{R}^{\intercal}$. Then, for trait distribution π_x, the variance in $f(X, Y)$ is given by $\sigma^2 = \mathbb{E}_{X,Y}[f(X, Y)^2]$ which equals the double integral, $\int_\Omega \int_\Omega f(x, y)^2 \pi_x(y)\pi_x(x)dydx$. Then, substituting into Equa-

[27] Note that $\rho = 1/2$ guarantee perfect transitivity but $\rho = 0$ does not guarantee that the tournament is perfectly cyclic. A counterexample suffices to explain why. Suppose each competitor chooses rock, paper, or scissors uniformly and independently. Suppose there are three competitors and the tournament is complete. Then, in order for the tournament to be perfectly cyclic, rock must be chosen by one competitor, scissors by another, and paper by the last. There are 6 ways this can happen but there are 27 possible tournaments. Therefore a three competitor system has a 21/27 chance of being perfectly transitive, even when the underlying performance function is clearly cyclic.

tion (4.49):

$$\rho = \frac{\int_\Omega \left(\int_\Omega f(x,y)\pi_x(y)dy\right)^2 \pi_x(x)dx}{\int_\Omega \int_\Omega f(x,y)^2 \pi_x(y)\pi_x(x)dydx}. \tag{4.51}$$

Note that the correlation coefficient is only large if it is possible to find some set of traits which are expected to perform either well or poorly on average, and if these traits occur with sufficient probability. That is, there must be some x such that $|\mathbb{E}_Y[f(x,Y)]|$ is large, and such that $\pi_x(x)$ is not too small. From this expression, it is not surprising that the expected strength of transitive competition is monotonically increasing in ρ. If there is a set of traits x which, on average, either overperform or underperform against randomly drawn opponents, and are frequently sampled, then a random sample of m competitors is expected to include some who perform well, and some poorly, against their neighbors. If, on the other hand, the expected performance conditioned on traits x is close to neutral, then ρ is small and competition is expected to be cyclic. In a rock-paper-scissors style game in which competitors are randomly and uniformly assigned rock, paper, or scissors, then conditioning on receiving a particular trait does not change the probability that an individual with that trait will win most contests, hence the tournament is expected to be highly cyclic.

Another way to read eq. (4.51) is as follows. Define the expected performance of traits x to be $\mathbb{E}_Y[f(x,Y)]$. Then since $\mathbb{E}_X[\mathbb{E}_Y[f(X,Y)]] = \mathbb{E}_{X,Y}[f(X,Y)] = 0$, $\mathbb{E}_X[\mathbb{E}_Y[f(X,Y)]^2]$ is the variance in the expected performance. Therefore ρ is the ratio of the variance in the expected performance to the variance in performance. A large variance in the expected performance means we are likely to sample some competitors who perform well, or poorly, against most opponents. Consequently, the sampled edge flow is expected to be more transitive than cyclic.

Rereading theorem 20 in this way leads to the following insight:

Corollary 20.1. *If the traits W, X, Y are sampled independently from π_x and $F = f(X, Y)$ then the correlation coefficient ρ is proportional to the variance in the expected performance:*

$$\rho = \frac{1}{\sigma^2}\text{cov}(f(X,Y), f(X,W)) = \frac{1}{\sigma^2}\text{Var}\left(\mathbb{E}[F|X]\right). \tag{4.52}$$

Let ν be the expected variance in the performance:

$$\nu = \frac{1}{\sigma^2}\mathbb{E}\left[\text{Var}(F|X)\right]. \tag{4.53}$$

Then $\nu = 1 - \rho$, so $\mathbb{E}[||F_c||^2]$ is monotonically increasing in ν, $\mathbb{E}[||F_t||^2]$ is monotonically decreasing in ν, and $\nu = \frac{1}{\sigma^2}\text{Var}[f(X,Y) - f(X,W)]$.

Proof. The proof of equation eq. (4.52) is given by equation eq. (4.51), and the fact that $\mathbb{E}[F] = 0$. Then $\nu = 1 - \rho$ follows by the law of total variance:

$$\sigma^2 = \text{Var}(F) = \mathbb{E}\left[\text{Var}(F|X)\right] + \text{Var}\left[\mathbb{E}(F|X)\right] = \sigma^2(\rho + \nu). \tag{4.54}$$

Since $\mathbb{E}[||F_c||^2]$ is decreasing in ρ, it is increasing in ν. Similarly, since $\mathbb{E}[||F_t||^2]$ is increasing in ρ, it is decreasing in ν.

The final expression for ν follows from $\sigma^2\nu = \sigma^2(1-\rho)$ which equals $\text{Var}[f(X,Y)] - \text{cov}[f(X,Y), f(X,W)]$. Since Y and W are i.i.d., $\text{Var}[f(X,Y)] = \frac{1}{2}(\text{Var}[f(X,Y)] + \text{Var}[f(X,W)])$. Substituting in gives $\sigma^2\nu = \frac{1}{2}\mathbb{E}[(f(X,Y) - f(X,W))^2]$. Since $\mathbb{E}[f(X,Y)]$ equals $\mathbb{E}[f(X,W)]$ this raw second moment is the variance in $f(X,Y) - f(X,W)$. □

Theorem 20 identifies which statistical feature of the trait distribution and performance function promotes transitive and suppresses cyclic competition. Corollary 20.1 complements this understanding by showing which feature suppresses transitive and promotes

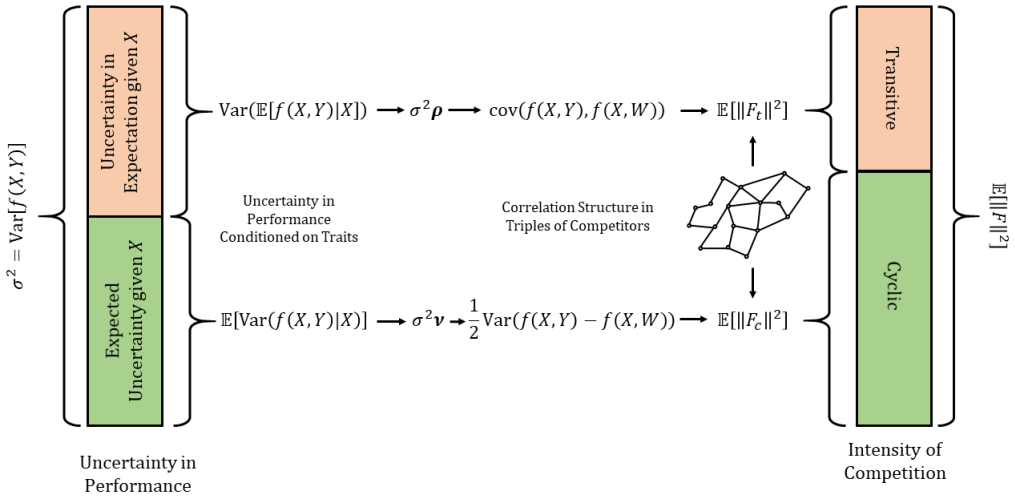

Figure 4.7: A schematic representing the conclusions of Theorem 20 and Corollary 20.1. The left hand side decomposes the uncertainty in performance into the uncertainty in the expected performance given X, and the expected uncertainty in the performance given X. These uncertainties are converted into ρ and ν which describe the correlation structure of triples of competitors. The sizes of ρ and ν, plus the topology of the network, determine the expected sizes of the transitive and cyclic components. Thus we convert a decomposition of the uncertainty in the performance into a decomposition of the intensity of the edge flow representing competition.

cyclic competition. Transitive competition is promoted by the uncertainty in expected performance, $\mathrm{Var}[\mathbb{E}(F|X)]$, and suppressed by the expected uncertainty, $\mathbb{E}[\mathrm{Var}(F|X)]$. Conversely, cyclic competition is suppressed by uncertainty in the expected performance, and promoted by expected uncertainty. If the uncertainty in expected performance is large, then we are likely to sample some competitors who are consistently better, or worse, than their neighbors, hence competition is mostly transitive. If the expected uncertainty in performance is large, then it is difficult to predict the performance of a single competitor against their neighbors, since performance is competitor dependent, hence competition is mostly cyclic.

Together Theorem 20 and corollary 20.1 provide conceptual bridges between uncertainty in the flow on each edge, correlation structure on edges that share an endpoint, and

cyclic/transitive structure on the network (see section 4.6.2). They establish the intuitive statements that conclude the introduction (p. 194). For example, the expected uncertainty in the performance of A against a random competitor is $\sigma^2\nu = \frac{1}{2}\mathbb{E}_X[\text{Var}_Y(f(X,Y)|X)]$. Thus, *"the less predictable the performance of A against a randomly drawn competitor, the more cyclic the tournament"* (see 1b). Then, by the equivalence of $\mathbb{E}_X[\text{Var}_Y(f(X,Y)|X)]$ to $\text{Var}(f(X,Y) - f(X,W))$, *"the more the performance of A depends on their opponent, the more cyclic the tournament."*

It remains to understand how the choice of trait dimension, trait distribution, and performance function influence ρ, and consequently the expected degree of cyclic competition. We provide an illustrative example below.

4.7 Example

Suppose that each competitor has a set of T traits. Assume that the traits are chosen so that the performance function $f(x,y)$ is non-decreasing in x_j, and non-increasing in y_j, for all j. This amounts to choosing a sign convention for each trait so that increasing any trait improves performance. Then a competitor with traits x has an advantage (in trait j) over an opponent with traits y if $x_j > y_j$.

In some events, competitors with a large advantage in a given trait can dominate, so that the event is primarily mediated by that trait. That is, competitors press their advantages. For example, a performance function of this type is the extremal performance function $f(x,y) = x_j - y_j$, where j is the dimension in which this difference is largest in magnitude, $j = \text{argmax}_j|x_j - y_j|$. In the extremal performance model, the performance is completely controlled by the largest advantage, so competitive events are as one-sided as possible, given the competitor's traits.

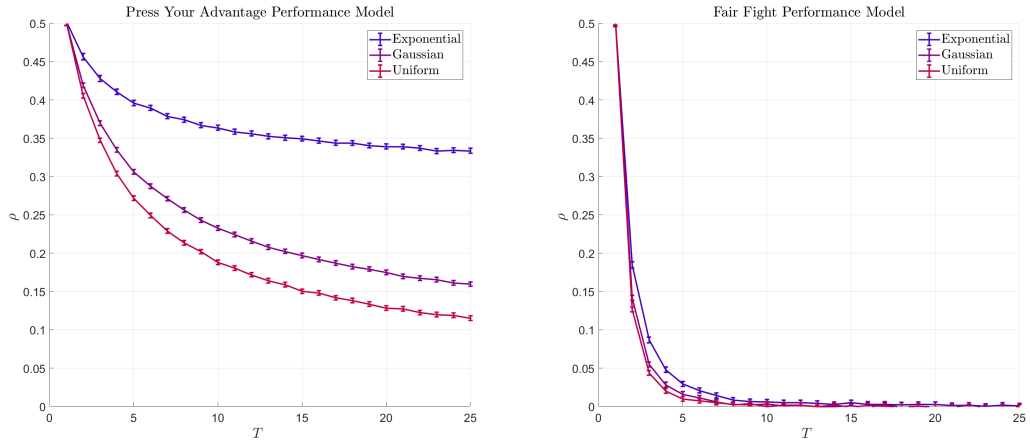

Figure 4.8: The correlation coefficient ρ for two different performance functions and three different trait distributions as a function of the number of competitive traits. Error bars represent three standard deviations in the estimated correlation coefficient. The “Press Your Advantage" panel shows $\rho(T)$ for the extremal performance model: $f(x, y) = x_j - y_j$ for j that maximizes the difference. The“Fair Fight" panel shows $\rho(T)$ for $f(x, y) = x_j - y_j$ for j that minimizes the difference.

Consider, in contrast, a competitive event in which competitors cannot press their advantages. For example: $f(x, y) = x_j - y_j$ for the dimension $j = \text{argmin}_j |x_j - y_j|$ that minimizes the advantage. This rule could model a contest in which competitors are required to reach a consensus about how to compete in advance or, where the weaker competitor controls which traits primarily mediate the competitive event. Competitors could be motivated or compelled to compete without pressing advantages by an external mediating body. For example, a sports league is motivated to keep teams evenly matched, even if the individual teams are motivated to win.

Suppose that the traits are drawn i.i.d from either an exponential, Gaussian, or uniform distribution. In each case, the variance of the trait distribution has no effect on ρ so, without loss of generality, each distribution is chosen to have variance one.

We estimated the correlation coefficient ρ for all six models (two performance functions, three distributions) with trait dimension varying from 1 to 25. To estimate the

correlation coefficient for a given model and trait dimension we sampled 10^6 triples of trait vectors X, Y, W and computed $f(X, Y)f(X, W)$. Averaging over all 10^6 triples gave an empirical estimate for the covariance, which was then normalized by an empirical estimate of the variance σ^2. Figure 4.8 shows the results.

For all three choices of trait distribution, $\rho(T)$ was larger if the extremal advantage model was used instead of the fair-fight model. This indicates that, the more competitors can press their advantages, the more transitive competition is, on average. This is not surprising, since in the fair-fight model, the traits mediating performance for competitor A against competitor B are likely different from the traits mediating competition between A and C. As a result, the success of competitor A is highly competitor dependent. Thus competition is more cyclic.

Note that this conclusion is much easier to test using the trait-performance theorem (Theorem 20) than by sampling a series of random edge flows. Using Theorem 20, we only needed to sample trait vectors for triples of competitors to evaluate ρ. This simplification greatly reduces the sampling cost.

In all six models tested, $\rho(T)$ is decreasing in T, so the expected proportion of competition that is cyclic is increasing. This matches the results in [117], where increasing the trait dimension typically decreased the expected degree of transitivity. This is intuitive, since larger T allows more ways for two competitors to compete, so it is harder to assign a single rating to a competitor.[28]

When using the extremal performance model the correlation $\rho(T)$ decays much faster in T for Gaussian and uniform traits than for exponential traits. This is because exponentially sampled traits are more likely to include large outliers. Since the extremal performance model sets f to the largest trait difference, the performance is more likely to depend on the

[28]Note that while this is often true it is *not* true for all trait-performance models.

outlier traits of each competitor. If a competitor has one particularly large trait, and T is large, then it is unlikely that any other competitor has a comparably large trait value in the same dimension. As a result, the competitor with the largest trait usually competes along that dimension and their performance against other competitors is fairly consistent. This leads to a relatively high ρ.

On the other hand, if the traits are drawn uniformly from $[0, 1]$ then no competitor can achieve a universal advantage by having one extremely large trait value. Instead, as the dimension of the trait space increases, competitors succeed by having a large trait value where their opponent has a small trait value - that is, by exploiting their opponents' weaknesses. In this situation, the relevant trait dimension that determines the outcome of competition depends on whom each competitor competes with. Consequently the correlation ρ becomes very small as T becomes large, so competition becomes predominantly cyclic.

In the fair-fight model all three trait distributions produce nearly identical correlations, since outlier traits do not mediate performance. Instead, performance is mediated by average traits, since the smallest advantage $X_j - Y_j$ is likely to come from a trait dimension where both X_j and Y_j are close to their expected values.

This example illustrates the explanatory power of the trait-performance theorem. By separating the influence of network topology from statistical assumptions about competition, the theorem facilitates numerical hypothesis testing and affords deeper insights by focusing the questions we ask about competitive tournaments.

4.8 Summary

The discrete HHD provides a natural, unified method for ranking and measuring intransitivity via a decomposition into perfectly transitive and cyclic components. The expected size of these components can be computed from the correlation structure of the edge flow. Using

a trait-performance model simplifies the correlation structure so that the decomposition of the edge flow can be related to the correlation in adjacent edges, and to a decomposition of the uncertainty in the edge flow. Intuitive statements about the expected sizes of the components can be rigorously proven for such models, which provide conceptual insight, as illustrated in Section 4.7. Future work should address other case studies, both inspired by real systems and chosen to illustrate generic behavior.

Further theoretical work could address random network topologies. If the network is sampled independently of the edge flow then the results of theorem 20 are largely unchanged. Future work might consider random networks whose distribution depends on the traits of each competitor, or ensembles whose traits are not i.i.d. For example, competitors who are neighbors in the network might have positively correlated traits. Correlation between network structure and traits would be important in an evolutionary setting where the hereditary nature of traits matters. Future work could also investigate null models with differently structured covariances in the edge flow. Studying these models will help contextualize the HHD when applied to real tournaments.

This work can is extended to data from real tournaments in Chapter 5. By studying win-loss records it is possible to infer the log-odds edge flow, and thus estimate the components of the HHD. Chapter 4 provides essential context by offering comparison to null models. Moreover, when an exhaustive win-loss record is not available, Theorem 20 suggests that the expected size of the cyclic component could be estimated by estimating the correlation coefficient ρ, which may be easier to estimate robustly.

Chapter 5

Application to Tournaments: Data

5.1 Preface

Like Chapter 4, this chapter is an expanded version of a manuscript that is in preparation for submission. The chapter is written so that it can be read independently of the other chapters, though the reader will gain a deeper appreciation for the approach and objectives if they read Chapter 4 first.

This chapter builds on the theory developed in Chapter 4. In order to apply the HHD to competitive tournaments directly the win probabilities must be known. These are rarely known in practice, so must be estimated from observed data. In this chapter we develop a series of estimation and hypothesis testing methods that are used to apply the HHD to data from real-world tournaments. Examples from politics and animal competition are explored (see Section 5.5 and Section 5.6).

5.2 Introduction

A tournament consists of a set of competitors who compete in pairwise competition events. Tournaments are familiar in sports, but also appear widely in animal behavior and politics. In virtually all tournaments there is great interest in ranking the competitors from best to worst. In sports, rankings are widely published, dictate draft orders and post-season schedules, and are obsessed over by fans. The famous school-yard question, who is the "Greatest of All Time" [149], is a ranking question. Outside of sports, rankings play an important role in decision making and data science. Examples in this vein include ranking colleges [154, 155, 156], ranking of web search results [124, 157], and ranking of movie suggestions on streaming platforms [158, 159, 160]. Some election systems can also be considered a ranking process. In biology ranking competitors is important as many animal societies are hierarchical, and dominance in a hierarchy is associated with greater reproductive success [119]. Thus ranking plays an essential role in our understanding of competitive systems across applications. In part, this ubiquity is due to rankings' attractive simplification of competitive systems into an ordered list.

Ranking is so pervasive that systems that cannot be consistently ranked due to rock-paper-scissor type cycles are almost universally treated as either surprising or disturbing. In psychology and economics, individuals with cyclic preferences are considered flatly "irrational", and the assumption that individuals have well ranked preferences is a founding axiom of choice theory [161]. In social choice theory and political science, cyclic preferences in the aggregate electoral opinion are considered "irrational", "paradoxical," and even "chaotic" [162, 163]. In biology, cyclic competitive systems are treated with somewhat less alarm, but with equal surprise and interest. Extensive theoretical work suggests that cyclic competition may maintain biodiversity by preventing competitive exclusion

[94, 99, 101, 105, 106, 107, 108], and may lead to deeply counter-intuitive evolutionary dynamics such as “survival of the weakest" [151]. As a result, there is an entrenched debate between two perspectives on competitive systems across fields. One camp, typically empirically motivated, argues that most tournaments are amenable to ranking and do not exhibit cyclic behavior. The other, supported by theory and select case-studies, argues that not all tournaments can be ranked, cycles play an important role in some systems, and that the ranking perspective is an impoverished simplification of real systems.

The debate described above is usually presented as a discussion of transitivity (linearity in some fields). A tournament is *transitive* if knowing A usually beats B, and B usually beats C, implies A usually beats C. If a tournament is transitive then there exists a unique ranking of the competitors that is consistent with the expected outcome between each pair of competitors. Not all tournaments are transitive, nor is it always clear from observed data whether a tournament is transitive [164]. Intransitive tournaments contain rock-paper-scissor type cycles in which A beats B beats C beats A, so are characterized by cyclic structure, not hierarchical structure. Two illustrative examples of observed cycles in competitive systems are presented below, one from politics and one from sports.

First we consider public opinion regarding intervention in Iraq’s invasion of Kuwait in 1990. Gaubatz [165] reconstructed public preferences for four different options of the U.S. response: withdrawal without involvement, sanctions, unilateral military intervention, or multilateral military intervention. Based on polling data Gaubatz concluded that aggregate public opinion was intransitive, with multiple preference cycles among the four options. According to Gaubatz 55 percent of the public preferred unliateral military intervention to multilateral military intervention, 69 percent preferred sanctions to unliateral military intervention, and 57 percent preferred multilateral military intervention to sanctions [165].

Second we consider the Houston Astros, Seattle Mariners, and Pittsburgh Pirates of American Major League Baseball. In 2019 the Astros had the highest win percentage of any baseball team in the American League (AL) West division (107 wins to 55 losses). In contrast the Mariners had the lowest win percentage in the AL West (68 to 94), and the Pirates had the lowest win percentage in their division, the National League Central, (69 to 93). Predictably, the Astros won 18 of the 19 games they played against the Mariners, with an average lead of 3.21 runs over the 19 games. The Mariners won all three of the games they played against the Pirates, with an average lead of 3.33 runs over the three games. How did the Pirates fair against the Astros? Seemingly against all odds the Pirates won two of the three games they played against the Astros, leading by an average of 6 runs over the three games - more than either the Astros led the Mariners or the Mariners led the Pirates. These sorts of underdog victories are not uncommon in baseball. Since 1980 fifteen to twenty percent of all triples of baseball teams have produced intransitive run records.

These examples raise two important issues regarding the transitivity/intransitivity debate. First, examples of cycles in competitive systems are often anecdotal and based on individual case-studies (cf. [102, 109, 103] or [166, 167, 168, 169]). While useful for illustrating that cycles can and do occur in important situations, case-studies cannot be used to study how frequently cycles occur [170]. Moreover, methodology differs between case-studies, making comparison difficult, and clouding the statistical significance of the collection of observed cycles [171]. In some cases, examples of dubious statistical significance are reported (cf. [172]), and in others, usually historical examples, statistical significance is only cursorily addressed if at all (cf. [168]). Therefore it is essential to move past documenting individual cases, and to perform repeatable analysis across many data sets simultaneously. Efforts to perform meta-studies of this kind have been performed

to different extents within each field (cf. [173, 112, 174, 175, 176]). Building on the meta-analysis approach we present a unified method that can be applied to commonly accessible data. The methods are fully documented in the appendix and publicly available code is provided online which implements the methods. To avoid relying on case studies that are difficult to compare we apply these methods to large scale datasets that allow for repeated analysis of systems across years, and comparison between systems. We highlight a few case-studies within the data sets, but contextualize the case-studies by comparing them to multiple examples from comparable competitive systems.

The second important issue raised by the examples is that intransitive cycles may be observed by chance when finitely many events are observed, even if the underlying system is transitive. This issue is raised clearly by the Pirates example. Are the intransitive run records observed in baseball flukes? Were the Pirates lucky in 2019, or is competition between some Major League Baseball teams inherently intransitive?

To distinguish these two scenarios we say that intransitivity is *structural* if A is expected to beat B is expected to beat C is expected to beat A, whereas intransitivity is *incidental* if A happened to beat B who happened to beat C who happened to beat A. Structural intransitivity is intrinsic to the win probabilities whereas incidental intransitivity is a consequence sampling error. Distinguishing between structural and incidental intransitivity is inherently a question of statistical significance as event data can only tell us what happened, not the probability of what happened. The standard approach is to propose a measure of intransitivity/transitivity that can be used as a test statistic. The measure is evaluated on the data, and is compared to the distribution of values it could take on under a null hypothesis. If the measure is much larger or smaller than what is expected then the null hypothesis is rejected. Multiple measures of intransitivity/transitivity have been proposed and are used as test of transitivity [133, 143, 16, 97, 94, 104, 100, 114]. In this paper we will use

the measure proposed by Jiang et al., Hodge intransitivity [16]. A critical shortcoming in some areas of the transitivity/intransitivity debate is that statistical significance of observed cycles is not always addressed [171], thus it is not always possible to tell whether observed cycles are structural or incidental. Jiang et al. do not address statistical significance in [16]. To avoid this issue we develop methods for quantifying the uncertainty in each quantity we estimate, and complement our estimation techniques with hypothesis testing.

Measuring intransitivity is especially important in systems that are structurally intransitive since the extent and intensity of cycles will determine how much is lost by trying to reduce the tournament description to a ranking. Structural cycles can alter long term dynamics [101, 105, 106, 107], can promote diversity amongst competitors [94, 99, 108], and can alter optimal strategies [18].[1] Alternatively, in a decision making context such as an election, the extent of cycles will determine how difficult it is to make a decision, and how arbitrary the ultimate decision is. This is the principle reason cycles are of such concern to social choice theorists. If aggregate preferences are highly cyclic then the outcome of elections can depend heavily on strategic deal-making [177], the order in which choices are presented [166], or the individuals with the agenda setting power [168]. At its most extreme, underlying cyclic preferences can lead to the Rikerian view of politics in which political outcomes are determined by strategic manipulation of factions not popular opinion [170, 169, 178]. These considerations strongly motivate the need for an understanding of both whether and to what degree real competitive systems are cyclic.

The goal of this chapter is to quantify how cyclic different competitive systems are, and

[1]The prevalence of cycles can reflect the range of successful strategies and degree to which competitor's must adapt their strategy to suit their opponent. If you know your opponent will choose rock you should choose paper, if you know your opponent will choose scissors you should choose rock, and if you know your opponent will choose paper you should choose scissors. If, on the other hand, competitive ability is described by a single number then the objective is simply to maximize that number, so the optimal strategy is opponent independent.

to identify cycles within those systems. The methods proposed are designed to be easily applied to different data sets so that they can be compared. The methods also incorporate uncertainty quantification and hypothesis testing in order to distinguish when observed results are likely structural or possibly incidental. All of the methods used in this chapter are documented in the appendix, and implemented in publicly available code. All of the data used has also been made publicly available. The data sets include some interesting and relevant cases (for example, the 2016 and 2020 American presidential elections), but are primarily chosen for their breadth of scope.

The chapter proceeds as follows. The basic theory needed to understand our methods is presented in Section 5.4. We emphasize the distinctions between our methods and standard methods for studying cyclic competition, and argue that our method provides a more nuanced understanding of competition than the classical categorization into the transitive and intransitive classes. Estimation, uncertainty quantification, and hypothesis testing methods are briefly summarized. In Sections 5.5 and 5.6 we apply our methods to a series of examples. The first example section considers a series of elections drawn from Danish, Dutch, and American politics. American presidential elections in 2000, 2016, and 2020 are highlighted as case-studies. The second example section considers a series of experiments in animal behavior regarding competition between birds in a pecking order. These examples are based on data from a meta-study of intransitivity in competition between animals [95]. Our choice of method is motivated separately for each of these two example applications in order to demonstrate the relevance of the analysis in each field and to highlight important similarities and differences between fields. Results from an analogous study of Major League Baseball are briefly discussed in Section 5.7 in order to highlight the importance of uncertainty quantification and testing for the statistical significance of conclusions. To conclude possible future work is proposed.

5.3 The HHD Reviewed

A tournament is conveniently represented with a competitive network. A competitive network has one vertex for each competitor, and a pair of directed edges connecting competitors who could compete. Let V be the number of competitors and E be the number of edges. The directed edge from B to A is weighted by the probability A beats B, which is denoted p_{AB}. The set of probabilities p are the win probabilities.

The Hodge-Helmholtz Decomposition (HHD) is a decomposition of an edge-flow on a graph [16]. In Chapter 4 we advocated for the log-odds edge flow when considering competitive tournaments. The log-odds edge flow is defined by first indexing all of the pairs of connected competitors, and assigning each edge an arbitrary reference orientation. If k indexes an edge let $i(k), j(k)$ denote the start and the end of the edge respectively. Then the log-odds edge flow is the vector f where $f_k = \log(p_{j(k)i(k)}/p_{i(k)j(k)}) = \text{logit}(p_{j(k)i(k)})$ is the log of the odds $j(k)$ beats $i(k)$. If $f_k > 0$ then $j(k)$ is expected to beat $i(k)$. If $f_k < 0$ then $j(k)$ is expected to lose to $i(k)$. When $f_k = 0$ the two competitors are equally matched. In general, large $|f_k|$ indicates that the competition event is predictable and the competitors are unevenly matched, while small $|f_k|$ indicates that the competition event is unpredictable and the competitors are evenly matched. Since $\text{logit}(p)$ diverges to $\pm\infty$ as p goes to zero or one we assume that none of the win probabilities equal zero and one.

The HHD decomposes f into two components, each drawn from a different class of tournaments.

The first class is the class of perfectly transitive, or arbitrage free, tournaments. A tournament is arbitrage free if the probability of observing a cyclic sequence of wins does not depend on the direction around the cycle (A beats B beats C beats A is just as likely as A beats C beats B beats A). A tournament is arbitrage free if the sum of f around any

cycle is zero.

If a tournament is arbitrage free then there exists a rating such that the edge flow on edge k is given by the difference in the ratings of the competitors at either end of the edge; $f_k = r_{j(k)} - r_{i(k)}$. This implies that $p_{AB} = \text{logistic}(r_A - r_B) = \exp(r_A)/(\exp(r_A) - \exp(r_B))$. The Elo rating method [137] assumes that the win probabilities satisfy the first form, while the Bradley-Terry rating system [140, 139] assumes that the win probabilities satisfy the second form for some ratings $\exp(r)$, which are widely used predictive rating systems [179, 137, 139, 140, 119, 138, 148, 118, 142]. A predictive rating system assumes that the probability A beats B can be expressed as a function of the ratings of A and B. In Chapter 4 we showed that any tournament that satisfies the Elo or Bradley-Terry model must also be arbitrage free. Thus a tournament satisfies the Elo or Bradley-Terry models if and only if it is arbitrage free.

If a tournament is arbitrage free then it is necessarily transitive. In a transitive tournament, knowing $p_{AB} > 1/2$ and $p_{BC} > 1/2$ implies $p_{CA} > 1/2$. Arbitrage free tournaments satisfy a stronger transitive property. If p_{AB} and p_{BC} are known then p_{CA} can be computed from p_{AB} and p_{BC}.[2] This is true for any edge in a cycle if the win probabilities on the other edges are known. Thus the value of the win probability, not just the sign relative to $1/2$, on any edge in a cycle is determined by the win probabilities on the other edges. For these reasons we refer to the class of arbitrage free tournaments as perfectly transitive.

In contrast, we define the class of favorite-free tournaments. A tournament is favorite free if the probability a competitor beats all of their neighbors in a row is equal to the probability that they lose to all of their neighbors in a row. A tournament is favorite free if the sum of f over each neighborhood equals zero. All favorite free tournaments are cyclic.

[2]Further, if $p_{AB} > 1/2$ and $p_{BC} > 1/2$ then $p_{AC} > \max\{p_{AB}, p_{NB}\} > 1/2$. The inequality $p_{AC} > 1/2$ is sometimes called weak stochastic transitivity, and the inequality $p_{AC} > \max\{p_{AB}, p_{NB}\}$ is sometimes called strong stochastic transitivity [180, 164].

A tournament is cyclic if for every path from A to B, such that every edge in the path is crossed from the expected loser to the expected winner, then there exists a path from B to A with the same property. Thus every path segment can be completed to form a cycle.

Like arbitrage free tournaments, which are specified by a set of ratings, favorite free tournaments are specified by a set of vorticities. Each vorticity is associated with a loop in the network, and the value of the vorticity represents the tendency of the system to cycle around that loop. The probability of observing $j(k)$ beat $i(k)$ is a sum of all the vorticities on loops that include edge k. For these reasons we refer to the class of favorite free tournaments as perfectly cyclic.

The HHD uniquely decomposes an edge flow f into a perfectly transitive edge flow f_t and a perfectly cyclic edge flow f_c [16]. To perform the HHD, we define the discrete gradient operator $G \in \mathbb{R}^{E,V}$ to be the matrix the incidence matrix of the directed graph given by introducing a directed edge from $i(k)$ to $j(k)$ for each k. Then:

$$[Gr]_k = r_{j(k)} - r_{i(k)}. \tag{5.1}$$

the space of perfectly transitive tournaments is the space of tournaments whose log-odds edge flow is in the range of G. The space of perfectly cyclic tournaments is the space of tournaments with edge flow in the null space of $G^\intercal$. Therefore there exist unique f_t and f_c such that:

$$f = f_t + f_c \text{ such that } f_t \in \text{range}\{G\}, f_c \in \text{null}\{G^\intercal\}. \tag{5.2}$$

where f_t is the orthogonal projection of f onto the range of G, and f_c is the orthogonal projection of f onto the nullspace of $G^\intercal$. The components can be solved for by first solving

the associated least squares problem for the ratings:

$$r = \text{argmin}_{u|\sum_{i=1}^{v} u_i}\{||Gu - f||^2\}. \tag{5.3}$$

This is the same least squares problem used in an unweighted log-least squares rating system. Least squares rating systems and log-least squares systems are widely used in pairwise comparison and sports rankings [128, 126, 127, 123, 130]. The transitive component is given by setting $f_t = Gr$, and the cyclic component is given by setting $f_c = f - f_t$. Therefore, if f is known then all that is required to perform the HHD is to solve the least squares problem defined by Equation (5.3). This least squares problem can be solved efficiently, even for large networks, since the gradient operator is sparse, and gets sparser the more competitors are added to the network [16].

The components of the HHD can be used to measure how cyclic a tournament is. The natural measures in this context are $||f_t||_2$ and $||f_c||_2$. These are the absolute sizes of the transitive and cyclic components. The Hodge intransitivity measure is the size of the cyclic component [16]. The absolute size of the cyclic component is the distance of f from the perfectly transitive subspace, and is the minimum amount of error needed when attempting to approximate f with the gradient of some rating r. Therefore $||f_c||_2$ is a measure of how far the given tournament is from satisfying the assumptions underlying either the Elo or Bradley-Terry rating systems. By comparing the sizes of the two components we can determine whether competition is principally transitive or principally cyclic. This motivates the relative measure: $||f_c||_2/||f||_2$. The relative measure is zero if competition is perfectly transitive and one if competition is perfectly cyclic.

We choose to use the Hodge intransitivity measure instead of other existing measures (cf. [97, 117, 100, 95, 104, 114]) for two conceptual reasons:

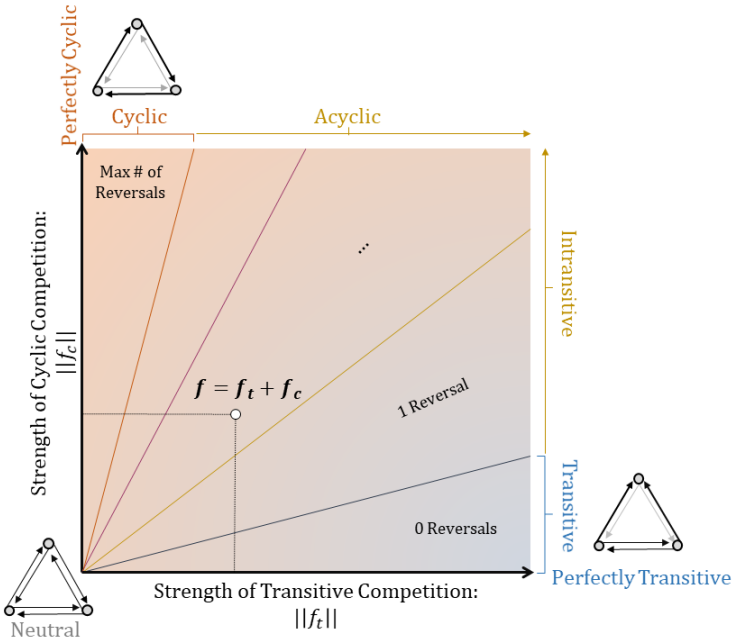

Figure 5.1: Conceptual representation of an arbitrary tournament according to the sizes of the transitive and cyclic components of the edge flow. The horizontal axis corresponds to the size of the transitive component, and the vertical axis to the size of the cyclic component. At the origin all win probabilities equal $1/2$ so the tournament is neutral. As either component becomes large the win probabilities approach either 0 or 1. If there is no transitive component then the tournament is perfectly cyclic, if there is no cyclic component then the tournament is perfectly transitive. Classification into the transitive and intransitive, or cyclic and acyclic classes depends on the sign of the edge flow, thus the level sets of discrete measures of transitivity/intransitivity correspond to angular sectors in the plane.

1. The Hodge measure is really one of a pair of measures. These two measures are the sizes of the two components of the decomposition. Thus the HHD provides a two-dimensional characterization of competitive systems rather than a one-dimensional characterization. From this perspective, standard intransitivity measures evaluate of the relative, not absolute, sizes of the components. By using the HHD, we can study cyclic competition separate from transitive competition, rather than as the opposite end of the transitive-intransitive spectrum. This separation allows for a more nuanced

analysis. Following this line of thought we propose a graphical approach. Let the size of each component be represented by two perpendicular axes. Then any tournament can be represented as a point in the positive quadrant of a plane. We will use this plane to graphically represent each tournament studied. We will then show that tournaments drawn from the same system typically cluster, and that the clusters corresponding to different systems are distinct, even if all of the systems are transitive. This is strong evidence that the HHD successfully characterizes properties of tournaments that are shared within a system but differ between systems. This is illustrated in Figure 5.1.

2. Unlike existing measures the Hodge measure is continuous in the win probabilities. To clarify the significance of this point consider two example systems. In the first system it is observed that A beat B 95 out of 100 games, and B beat C 95 out of 100 games. In the second system it is observed that A beat B 55 out of 100 games and B beat C 55 out of 100 games. How often do we expect A to beat C in the two systems? Assuming transitivity, we expect A to beat C in both examples. However, we might also reasonably expect that A would beat C more often in the first example than in the second example. This distinction reflects a stronger notion of transitivity. The former conclusion is based on the logic that observing that A usually beat B, and B usually beat C, implies A usually beats C. The latter conclusion is based on the logic that seeing the frequency with which A beat B, and the frequency with which B beat C, implies the frequency with which A beats C. The latter conclusion reflects stronger assumptions about the structure of the win probabilities. The former conclusion only depends on whether win probabilities are greater than or less than a half, while the latter depends on the values of the win probabilities. The Hodge measure reflects deviations from this stronger set of

assumptions, namely, that there exists an Elo/Bradley-Terry rating of the competitors which predicts the win probabilities. Thus, if it were observed that A actually lost to C 40 out of 100 games the Hodge measure would return a larger value in the former case than the latter. Even if A beat C 51 out of 100 games, the Hodge measure would not be zero, despite the fact that the observed outcomes are transitive. In both cases we would expect A to beat C at least more than 55 out of 100 games, so the Hodge measure would be nonzero since C won surprisingly often. This latter expectation, that $p_{AC} > \max\{p_{AB}, p_{BA}\}$ if $p_{AB}, p_{BC} > 1/2$ is sometimes refered to as strong transitivity [180, 164]. Figure 5.2 illustrates the three different transitivity assumptions. Since many, if not most, observed empirical systems are plausibly transitive [98, 112, 176], it is useful to have a measure which can detect violations of a stronger hypothesis, and thereby detect a latent cyclic component. The fact that the Hodge measure is continuous also makes the Hodge measure less prone to sampling error when observed event counts are near 50-50. If A beat C 51 out of 100 games reversing only two outcomes between A and C would change the system from transitive to intransitive, completely reversing the system's classification. Thus the classification into transitive and intransitive is sensitive to small sampling errors when win counts are near a half. In contrast, the Hodge measure would barely change when the two outcomes are reversed. Because the Hodge measure is continuous in the win probabilities small changes in sampled outcomes never lead to disproportionately large changes in conclusions.

The measures produced by the HHD can also be related to statistical properties of the tournament. In Chapter 4 we demonstrated that, if competition is modelled using a trait-performance model, then the expected sizes of the components squared can be computed explicitly from the dimensions of the network and some simple statistics. In

Intransitive:

$$P = \begin{bmatrix} . & 0.8 & 0.8 & 0.2 \\ 0.2 & . & 0.8 & 0.8 \\ 0.2 & 0.2 & . & 0.8 \\ 0.8 & 0.2 & 0.2 & . \end{bmatrix}$$

Transitive but not Strongly Transitive:

$$P = \begin{bmatrix} . & 0.8 & 0.8 & 0.6 \\ 0.2 & . & 0.8 & 0.8 \\ 0.2 & 0.2 & . & 0.8 \\ 0.4 & 0.2 & 0.2 & . \end{bmatrix}$$

Strongly Transitive but not Perfectly Transitive:

$$P = \begin{bmatrix} . & 0.8 & 0.8 & 0.9 \\ 0.2 & . & 0.8 & 0.8 \\ 0.2 & 0.2 & . & 0.8 \\ 0.1 & 0.2 & 0.2 & . \end{bmatrix}$$

Perfectly Transitive:

$$P = \begin{bmatrix} . & 0.800 & 0.941 & 0.996 \\ 0.200 & . & 0.800 & 0.941 \\ 0.059 & 0.200 & . & 0.800 \\ 0.0004 & 0.059 & 0.200 & . \end{bmatrix}$$

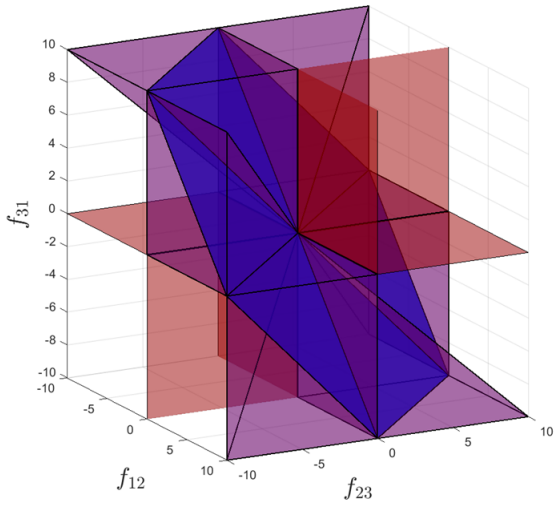

Figure 5.2: The left panel provides four different matrices containing win probabilities, where the ij entry is the probability i beats j. The first is intransitive because of the cycle $1 > 2 > 3 > 4 > 1$. In general, if there is no way to reorder the indices such that all entries above the diagonal are greater than 0.5 then the system is intransitive. The next is transitive, because it can be ordered so that all entries above the diagonal are greater than 0.5, however it is not strongly transitive. To be strongly transitive the entries across each row must be increasing as read from right to left, the entries along each column must be decreasing as read top to bottom. This is violated by the entry in the top right corner. The next example is strongly transitive, but not perfectly transitive, since the win probabilities do not match an Elo type predictive rating for any ratings. The right hand panel shows the regions corresponding to each class of tournament for a system with three competitors. Each axis represents the log-odds. The octants bounded in red are cyclic. Any set of log-odds not contained in the red octants are transitive. The purple shaded region is the space of log-odds that are strongly transitive, and the blue shaded subspace is the space of log-odds that are perfectly transitive. Note that the each stronger notion of transitivity is nested inside of each weaker notion.

a trait-performance model it is assumed that the probability that one competitor beats another can be expressed as a function of the traits of the two competitors, and that all competitors' traits are drawn i.i.d. from a trait distribution that models the demographics of the competitor pool. Then the ratio of the expected size of the cyclic component (squared) to the expected size of the edge flow (squared) only depends on the density of the network,

and the correlation, ρ, in the log-odds that A beats B and A beats C. The denser the network the larger the ratio, and the larger ρ the smaller the ratio. Therefore, increasing the correlation in the performance of A against B with A against C promotes transitive competition, while decreasing the correlation promotes cyclic competition.

5.4 Estimation Methods, Uncertainty Quantification, and Hypothesis Testing

5.4.1 Bayesian Methods

Point Estimation

In practice, the win probabilities are rarely known, so the log-odds, f, must be estimated from data. Suppose that the outcomes of a series of competition events are recorded. Let $n \in \mathbb{Z}^E$ record the number of events observed on each edge, not including ties. Let $w \in \mathbb{Z}^E$ be the number of wins observed on each edge, where w_k is the number of times $i(k)$ beat $j(k)$. Then our objective is to estimate f given n and w. The following section summarizes the key results. For details see Appendix A

To help constrain the estimates we assume that the win probabilities are distributed according to a symmetric beta distribution with parameters $\alpha, \beta = \gamma > 0$. The beta distribution is the conjugate prior for this estimation problem and is widely used to estimate binomial parameters [164, 181, 182, 183]. If $\gamma = 1$ the prior is uniform, if $\gamma < 1$ then the prior is large for win probabilities far from $1/2$, and if $\gamma > 1$ then the prior is large for probabilities near $1/2$. Introducing the prior with parameter γ is equivalent to not introducing a prior and adding a fictitious $\gamma - 1$ wins and $\gamma - 1$ losses to each edge. The prior parameter can be chosen according to existing standards [181, 182, 183], or can be fit

for. Details on maximum likelihood estimation of γ are provided in the appendix. When available, the prior parameter γ should be fit to win-loss data from a tournament that is not the tournament of interest, but is expected to have the same underlying statistics. For example, if the tournament of interest is a particular baseball season, then past baseball seasons could be used to fit for γ.

Under these assumptions the estimation problem is equivalent to logistic regression, where the outcomes are either a win or a loss, and the predictor variable is the indicator vector e_k for which pair of competitors are competing.

Given n, w, γ the win probabilities P_k are beta distributed with $P_k \sim \text{Beta}(w_k+\gamma, n_k - w_k + \gamma)$. The conditional expectation for P_k given n_k, w_k, γ is $\mathbb{E}[P_k|n_k, w_k, \gamma] = (w_k + \gamma)/(n_k + 2\gamma)$. This is the win frequency of $j(k)$ against $i(k)$ if γ wins and losses are added to the record.

Then, the log-odds that $j(k)$ beats $i(k)$, F_k, has posterior proportional to $\text{logistic}(f)^{w_k+\gamma}$ times $\text{logistic}(-f)^{n_k-w_k+\gamma}$, where $\text{logistic}(f) = (1+\exp(-f))$ is the inverse of the logit function. This distribution is unimodal, and its negative log is convex. The mode, or mean, of this distribution can be used as point estimators for the log-odds. The mode and mean are given by:

$$\begin{aligned} f_{\text{MAP}}(n_k, w_k, \gamma) &= \text{logit}(\mathbb{E}[P|n_k, w_k, \gamma]) = \ln\left(\frac{w_k+\gamma}{n_k+w_k+\gamma}\right) \\ f_{\text{exp}}(n_k, w_k, \gamma) &= \mathbb{E}[F_k|n_k, w_k, \gamma] = \psi(w_k+\gamma) - \psi(n_k - w_k + \gamma) \\ &= \sum_{w=0}^{w_k} \frac{1}{w+\gamma} - \sum_{l=0}^{n_k-w_k} \frac{1}{l+\gamma}. \end{aligned} \tag{5.4}$$

where $\psi(x)$ is the digamma function. These two estimators are asymptotically equivalent, and $f_{\text{exp}}(n_k, w_k, \gamma)$ converges to $f_{\text{MAP}}(n_k, w_k, \gamma)$ with discrepancy order $\min\{w_k, n_k - w_k\}$.

The conditional expectation, $f_{\text{exp}}(n_k, w_k, \gamma)$, can be easily updated after each observed

event. Suppose w_k wins have been observed and $n_k - w_k$ losses have been observed. If, on the next event, an additional win is observed then add $1/(w_k + \gamma)$ to the conditional expectation. If a loss is observed subtract $1/(n_k - w_k + \gamma)$ from the expectation. This updating scheme is self-correcting since surprising events lead to larger updates than events that are commonly observed.

The tails of the posterior distribution of F_k decay exponentially. As $f \to \infty$ the posterior decays with rate equal to the number of observed losses plus the number of fictitious losses, $n_k - w_k + \gamma$. Losses are evidence that F_k should be negative, thus the posterior distribution for F_k is constrained above by observing losses. Similarly, as $f \to -\infty$ the posterior decays with rate equal to the number of observed wins plus the number of fictitious losses $w_k + \gamma$. Wins are evidence that F_k should be positive, thus the posterior distribution for F_k is controlled below by observing wins. It follows that the distribution is skewed positive if more wins are observed than losses, and skewed negative if more losses are observed than wins. As a consequence the MAP estimator is generally more conservative than the conditional expectation.

The variance in the posterior is given by $\psi^{(1)}(w_k + \gamma) + \psi^{(1)}(n_k - w_k + \gamma)$ where $\psi^{(1)}$ is the trigamma function. The trigamma function is approximately one over its argument, thus the variance in the posterior is approximated by $1/(w_k + \gamma) + 1/(n_k - w_k + \gamma) + \mathcal{O}(\min(w_k, n_k - w_K)^{-2})$. It follows that the variance is only small if both the observed number of wins and losses is large. That said, because the tails decay exponentially at different rates the tail behavior of the posterior distribution is not well approximated by a Gaussian distribution, and the distribution may be highly skewed. For example, if ten wins and no losses are observed then the posterior will have a large variance since the distribution has a slowly decaying positive tail. Nevertheless the distribution has a rapidly decaying tail for $f < f_{\text{MAP}}$, so we can be confident that F is not much smaller than f_{MAP}, even if we

cannot be confident that F is not much larger than f_{MAP}. In order to answer some questions it may be enough to know that one of the tails of the posterior decays quickly. In that case the variance in the posterior can provide a misleading representation of the uncertainty in the posterior if used in isolation.

These observations about the posterior distribution can be used to introduce sample size requirements. If we wish to estimate the log-odds to within a desired variance then it is enough to check whether the variance in the posterior is small enough on each edge. This effectively requires $\min\{w_k + \gamma, n_k - w_k + \gamma\}$ to be larger than a threshold fixed by the desired maximum variance. Alternatively, if we only need to estimate the sign of the log-odds robustly then it is natural to put a lower bound on the distance from f_{MAP} to zero relative to the rate of decay of the inward tail, $\max\{w_k + \gamma, n_k - w_k + \gamma\}$. The desired bounds can be set relative to the size of the expected signal (MAP estimator) when sampled from the prior (see Appendix A). This technique could be used to choose the number of samples in an experimental setting where the sample size can be controlled, and should be fixed before the outcomes are observed.

To determine the accuracy and precision of these estimators suppose that the true log-odds, f, are known. Then n outcomes are observed, of which W are wins. Then either $f_{\text{MAP}}(n, W, \gamma)$ is computed or $f_{\text{exp}}(n, W, \gamma)$ is computed. Now the point estimators are random variables whose distribution depends on the true log-odds, number of samples observed, and prior parameter γ. The asymptotic accuracy of the estimator can be computed by evaluating the expected error in the estimator for large n. Similarly, the asymptotic precision of the estimator can be computed by evaluating the standard deviation in the estimator for large n. The expected error in the MAP estimator and conditional expectation estimator are both order n^{-1}. This bias arises from the prior, which encourages conservative estimates of the log-odds, and the curvature of the logit function, which

encourages overestimates of the log-odds. For $\gamma = 1/2$ the MAP estimator is unbiased to order n^{-2} and for $\gamma = 1$ the conditional expectation is unbiased to order n^{-2}. The prior parameter, γ, is greater than one for most of the case studies, so the point estimators are usually expected to be conservative, with biases order n^{-1}. The standard deviation in the estimators are both order $n^{-1/2}$. Thus both estimators are asymptotically unbiased, and for large n the bias in the estimators is smaller than the standard deviation. It follows that most of the error in the estimators is from uncertainty in W/n given a finite sampling size, not the bias in the estimators. An analysis of the asymptotic accuracy and precision of the estimators is provided in Appendix A.

The components of the HHD (rankings, transitive component, cyclic component) are all linear functions of the edge flow (see Equation (5.2)). It follows that the expected value of the components of the HHD are simply given by applying the HHD to $f_{\text{exp}}(n, W, \gamma)$. Then an estimated ranking is given by sorting the competitors according to their estimated rating. The variance in the components of the HHD can be computed directly from the variance in the posterior for each component as follows. If X is a random variable with covariance $\mathbb{V}$ and $Y = AX$ for some matrix A then the variance in Y is $A^{\intercal}\mathbb{V}A$. Since the variance $\mathbb{V}[F|n, W, \gamma]$ is known, the variance in the posterior distribution for the rankings, transitive flow, and cyclic flow can all be computed analytically.

Estimators for the measures can be introduced by evaluating the measures on the estimated flow. By evaluating the measures on the same estimated flow used to estimate the HHD components we maintain consistency across the analysis. An alternative option is to sample from the posterior distribution of flows, evaluate the measures for each sampled flows, and then form an empirical approximation to the posterior distribution for each measure. Sampling methods are discussed in Section 5.4.1. Given an empirical approximation to the posterior distribution of each measure it is easy to calculate the expected value and

uncertainty in the measure.

Both of these approaches to estimating the measures are biased. Let P_S be an orthogonal projector onto a subspace S. Then if F is a random edge flow with covariance $\mathbb{V}$, then the expected value of $||P_S F||^2$ is $||P_S \mathbb{E}[F]||^2$+trace$(P_S \mathbb{V})$ (see Section 4.6.2). Since $\mathbb{E}[F]$ typically converges to the true value of f in the limit of large sample size the principal source of error in this approximation is the second component, trace$(P_S \mathbb{V})$. This term is the contribution to the expected size of the measure (squared) due to uncertainty in F. In general, the more uncertainty there is in F, or the larger the dimension of S, the larger this term, and the larger the bias. The measures are defined by setting S to either the range of the gradient, or the null space of its transpose. These spaces have dimension equal to $V-1$ and $E-(V-1)$. The latter number is the cyclomatic number, which is the dimension of the cycle space of the network. If the graph is dense then $V-1 < E/2$, so the cyclic subspace is higher dimensional than the transitive subspace. Therefore, not only are both measures biased by uncertainty in F, the cyclic measure is typically more biased than the transitive measure. Thus it is easy to overestimate how intransitive a network is if the uncertainty in F is moderately large, and the difference in the dimensionality of the transitive and cyclic subspaces is not accounted for.

Here we advocate for estimating the measures by applying them directly to the estimated flow, not by approximating the expected value of the measures over their posterior. We advocate for this approach since it effectively halves the bias due to uncertainty (see Appendix A).

Let f denote the true, but unknown, value of the log-odds. Then a sample of W wins is observed out of n games. The estimators are functions of W that return a “best" estimate of f given the observed win record. If the measure is evaluated on this estimator for f then the only source of uncertainty is the uncertainty in W. Alternatively, given the

observed win record it is possible to compute the posterior distribution of F, and therefore, to approximate the posterior distribution for the measures. Averaging over this posterior distribution is equivalent to computing the expected value of the measures when evaluated on random edge flow F, where F is drawn from the posterior for the log-odds. The random edge flow F the same expected value as the point estimator $f_{\text{exp}}(n, W, \gamma)$. However, by the law of total variance, the variance in F is equal to the variance in $f_{\text{exp}}(n, W, \gamma)$ plus the expected value of the variance in the posterior when W is drawn randomly. Therefore the variance in F is strictly larger than the variance in the point estimator $f_{\text{exp}}(n, W, \gamma)$. In the limit of a large sample size the expected variance in the posterior converges to the variance in the expected F, thus the variance in F converges to twice the variance in the point estimators. Approximating the measure using its posterior effectively doubles the bias due to uncertainty.

The asymptotic equivalence of the variance in f_{exp} and the expected variance in the posterior means that we can estimate the size of the bias introduced to the measures due to uncertainty. Then it is possible to compute what percent of the estimated measure is expected to have come from uncertainty. This percent can be used as a benchmark for whether or not the estimated value is reliable.

In total the estimation procedure consists of the following steps. First, estimate the log-odds edge flow using Equation (5.4). Next, apply the HHD to the estimated flow and evaluate the measures on the estimated flow. Then, compute the variance in the posterior distribution for the flow on each edge. Using the variance in the posterior for the flow, compute the variance in the estimated components and the percent of the estimated measures that is expected to have come from uncertainty. Finally, compare the variances in each estimated component and expected biases to desired precision and accuracy benchmarks.

Sampling and Interval Estimation

Point estimation is only reliable if enough events have been observed on each edge so that the variance in the posterior distribution for the flow is small. The variance in the posterior for the flow can be used to compute the variance in the ratings, HHD components, and to approximate the uncertainty in the measures. This approach is limited in that it does not provide an uncertainty estimate for the rankings, nor does it provide confidence intervals on any of the estimated quantities. Credible intervals are of particular interest since the posterior for the flow is often highly skewed.

Sampling can be used to quantify the uncertainty in the estimated flow, HHD components, rankings, ratings, and measures. Sampling from the posterior is straightforward, as the win probabilities given observed win records are all beta distributed. To sample, draw $P_k \sim \text{Beta}(w_k + \gamma, n_k - w_k + \gamma)$, and set $F_k = \text{logit}(P_k)$. Then evaluate the HHD on the sampled flow, rank the competitors, and compute the value of the measures.

Sampling is particularly useful for quantifying the uncertainty in the rankings. The Spearman and Kendall rank correlation of the sampled rankings provide measures of how consistent the competitor ranking is across the posterior for the ratings.

Samples can also be used to approximate credible intervals for any of the desired quantities. All reported credible intervals are given by finding the highest posterior density interval (HPDI) [184] for a histogram generated by the samples. Thus the point estimators can be complemented by interval estimation.

Sampling can also be used to estimate the posterior probability that the tournament is transitive, and to estimate the posterior probability that the tournament is strongly transitive. Once a set of win probabilities have been sampled we can easily test for transitivity. If the tournament is complete then the Landau linearity index h (see Section 4.4) can be computed analytically from the variance in the number of competitors each competitor is expected to

beat (see 4.4.2) [3]. If $h = 1$ then the tournament is transitive. Alternatively, if the network is not complete then we can search for a ranking of the competitor's that is consistent with the sampled win probabilities. If such a ranking can be found then the network is transitive. To check whether the network is transitive we use an iterative method to gradually build a nominal ranking. At each step we update a nominal list of all individuals we expect to be lower and higher ranked than each other competitor based on the assumption the network is transitive. These lists are expanded one edge at a time until we have either found an inconsistency, or checked every edge. Individuals are added to each list based on the following method. Consider edge k. If $p_{i(k)j(k)} > 1/2$ then $i(k)$ must have a lower rating than $j(k)$. Therefore $j(k)$ is added to the list of individuals $i(k)$ dominates. Moreover, since we assume the network is transitive, all individuals dominated by $j(k)$ are added to the list of individuals dominated by $i(k)$, and all individuals who dominate $i(k)$. If at some point this process produces an inconsistency $i(k)$ dominates $j(k)$ and $j(k)$ dominates $i(k)$ then the network is not transitive. The fraction of sampled tournaments that are transitive approximates the posterior probability that the true tournament is transitive.

If the sampled tournament is transitive then we also check if it is strongly transitive. If the sampled tournament is transitive then we can organize the competitors in a rank order. This ranking is produced implicitly by our transitivity check. Once in rank order, a tournament is strongly transitive if the matrix of win probabilities is increasing across any row (read from left to right), and decreasing down any column (read top to bottom). This can be easily checked. The fraction of sampled tournaments that are strongly transitive approximates the posterior probability that the true tournament is strongly transitive.

[3] Incidentally, by computing h for each sample we compute an approximation to the posterior distribution of the linearity index.

Limitations

All of the methods described thus far have relied on a Bayesian approach to the estimation problem. The Bayesian approach requires a prior distribution over the space of possible models (edge flows). Without a prior the posterior distribution is not well defined. The less data is observed the more the posterior resembles the prior, and the more the associated estimates resemble their expectation under the prior. Therefore, when only a small number of events is observed, the results of the Bayesian approach depend heavily on the information provided by the prior.

Throughout we have assumed that the win probabilities are all sampled i.i.d. from a symmetric beta distribution with parameter γ that is fit to past events. The beta distribution was chosen since it is the conjugate prior to the binomial. We required that the prior is symmetric to ensure that the win probabilities are not biased by the choice of (arbitrary) edge orientations. The win probabilities were assumed to be drawn i.i.d. so that γ could be robustly estimated from past events, or from the tournament as a whole, and so that the prior captures a basic model for the type of competition, not a model for competition between a particular pair of competitors.

However, assuming independence of the win probabilities on the edges biases the estimators for the measures since the space of perfectly cyclic tournaments is higher dimensional than the space of perfectly transitive tournaments. If the win probabilities are sampled from the prior then, using Equation (4.37) , the expected value of the squared measures are $2(V-1)\psi^{(1)}(\gamma)$ and $2(E-V+1)\psi^{(1)}(\gamma)$. Therefore, when the network is dense, and the win probabilities are sampled from the prior, the resulting tournament is typically more cyclic than it is transitive. Hence, when there is limited data the Bayesian approach tends to return cyclic estimates. This bias is not an error, since under the chosen prior most models are more cyclic than transitive. Therefore, when there is uncertainty

in which model could have generated the observed data it is natural to give more cyclic answers as there most of the credible models are usually more cyclic than transitive.

This sort of bias could be avoided if a different prior was used to model the win probabilities. If the win probabilities are not assumed to be independent of each other then some correlation structure must be assumed on the win probabilities. The choice of correlation structure changes the expected sizes of each component when sampled from the prior, and thus the posterior (see Section 4.6). It follows that, unless a particular correlation structure is known ahead of time, and is a natural model for the type of competition observed, picking a correlation structure a priori also amounts to biasing our prior estimate of the sizes of the components. Since our goal is to compare the sizes of the components across diverse tournaments we cannot assume a correlation structure a priori without also implicitly making assumptions about the sizes of the components.

These biases are the principal challenge when using the Bayesian approach. The uncertainty in the edge flow is only small if enough wins *and* losses are observed between each pair of competitors. Otherwise the rare events lead to a slowly decaying tail. Then this uncertainty biases the estimated measures, but does not bias them equally, leading to a systematic tendency to overestimate the cyclic component relative to the transitive component.

As it is unusual to find data sets with many wins and losses between every pair of competitors it is important to complement the Bayesian approach with alternative statistical approaches. In the next section a set of frequentist approaches are introduced that aim to answer different but related questions. Instead of attempting to estimate the true values of the measures, which requires a large amount of data, these approaches propose and test hypotheses about the measures that require less data to test.

5.4.2 Frequentist Methods

An alternative approach to Bayesian estimation is to use a frequentist approach to test hypotheses regarding the sizes of the components. A hypothesis testing approach allows us to ask and answer different questions that require less data.

For example, suppose we observe A beat B ten out of ten games, B beat C ten out of ten games, and C beat A ten out of ten games. Then, since we have only observed wins on each edge the variance in the posterior for the edge flow is large, and the upper tails of the distribution decay slowly. This means it is not possible to provide a confident estimate of the value of the log-odds on each edge. Were the win probabilities 90 percent (log-odds 2.19), 99 percent (log-odds 4.60), or 99.9 percent (log-odds 6.91)? Moreover, since we cannot estimate the value of the log-odds we cannot estimate the size of the cyclic component. That said, we should be able to put a lower bound on the log-odds confidently, since the lower tail of the posterior is decaying quickly. These lower bounds should make clear that F is positive on each edge, thus competition is intransitive, with cyclic component much larger than the transitive component. So, even if we cannot robustly estimate the sizes of the components, we may be able to bound them.

This observation motivates the following two sections. First we develop tools to test the hypotheses:

1. H_t: The tournament is perfectly transitive.

2. H_c: The network is perfectly cyclic.

Then we extend these tools to give estimated lower bounds on the sizes of the components.

Hypothesis Testing

If f is a particular log-odds edge flow, then the log-likelihood of having observed w wins given n games can be computed analytically. Similarly, the maximum log-likelihood over all log-odds edge flows is given by setting the win probabilities p_k to the observed win frequency w_k/n_k, and then computing the probability of sampling $W_k = w_k$ wins. The corresponding log-odds, $f_k = \text{logit}(w_k/n_k)$, is the MLE edge flow. The corresponding log-likelihoods are given below:

$$\begin{aligned}
\ln(\mathcal{L}(f|w,n)) &= \sum_{k=1}^{E} \ln \binom{n_k}{w_k} - w_k \ln(1+\exp(-f_k)) - (n_k - w_k)\ln(1+\exp(f_k)). \\
\ln(\mathcal{L}(f_{\text{MLE}}|w,n)) &= \sum_{k=1}^{E} \ln \binom{n_k}{w_k} + w_k \ln\left(\frac{w_k}{n_k}\right) + (n_k - w_k)\ln\left(\frac{n_k - w_k}{n_k}\right)
\end{aligned} \tag{5.5}$$

Therefore the log-likelihood ratio is:

$$\begin{aligned}
\lambda(f|w,n) &= -2(\ln(\mathcal{L}(f|w,n)) - \ln(\mathcal{L}(f_{\text{MLE}}|w,n))) \\
&= 2\sum_{k=1}^{E} n_k D_{KL}\left(w_k/n_k || \text{logistic}(f_k)\right)
\end{aligned} \tag{5.6}$$

where $D_{KL}(p||q)$ is the KL divergence [185] between the distributions $p, 1-p$ and $q, 1-q$. Therefore the difference in the log-likelihoods is a weighted sum of the KL divergence between the observed win frequencies and the predicted win frequencies given model f. Edges with more events are weighted more heavily in the sum. Replacing the likelihood with the posterior in any of these expression simply requires adding γ to w_k and 2γ to n_k.

The test statistic $\lambda(f|w,n)$ is nonnegative, is large when there is a large discrepancy between the data and the proposed model f, and is asymptotically χ^2 distributed with E

degrees of freedom [186]. The likelihood ratio is chosen since it is the standard test statistic for logistic regression [187, 188]. The hypothesis $F = f$ is rejected if the test statistic $\lambda(f|w, n)$ is too large.

The lower boundary for rejection can be determined by computing the probability of sampling a win record W given win probabilities $p = \text{logit}(f)$, such that $\lambda(f|W, n) \geq \lambda(f|w, n)$. To approximate the probability of sampling a win record W with a larger test statistic we repeatedly sample W_k from the binomial distribution with n_k events and success probability $\text{logit}(f_k)$. For each sampled win record we compute the test statistic $\lambda(f|W, n)$. Then we compute the fraction of sampled win records which have a larger test statistic than $\lambda(f|w, n)$. If this fraction is less than a significance level α then the model is rejected. The number of samples drawn should be chosen based on the desired significance level. If the number of samples is chosen so that the standard deviation in the fraction of samples with larger test statistic is at least k times smaller than α then on the order of k^2/α samples should be drawn. We use $\alpha = 0.05$ throughout so the probability of false rejection is five percent. A smaller α could be chosen when more data is available.

Other test statistics can be used to evaluate the plausibility of the model f given the data. One natural choice is the Akaike Information Criterion (AIC) difference, between the proposed model, f, and the maximum likelihood model. This is equal to the log-likelihood difference λ minus $2E$ [189]. Subtracting off $2E$ accounts for the fact that the MLE estimate has E degrees of freedom that can be used to maximize the log-likelihood. The AIC is typically positive when the proposed model f is underfit, and negative when it is overfit. Using the AIC as a test statistic gives the same results as using the log-likelihood ratio since the two statistics only differ by a constant that depends on the hypothesis.

Another alternative is to use the log-likelihood of f itself as the test-statistic. This approach is appealing since it is equivalent to using the likelihood of the sample itself as

the test-statistic, and is thus equivalent to the standard p-test.

A final alternative is to start by computing a p-value for each edge. This is straightforward since the number of wins on each edge is binomially distributed given f and n. Then it is possible to compute the probability of sampling a number of wins that is equally or less likely than the actual number of wins observed on the edge. Performing this calculation produces a p-value for each edge. If f is the true log-odds then each of these p-values are independent samples from a uniform distribution. Thus the set of p-values should be distributed evenly between zero and one. To test whether the set of p-values could plausibly be E independent samples from a uniform distribution we use the Kolmogorov-Smirnov test [190, 191, 192]. The Kolmogorov-Smirnov test uses the supremum of the absolute difference between the empirical cdf corresponding to the observed p-values, and the cdf of the uniform distribution, as a test statistic. The set of p-values can be rejected if this distance is too large. The lower boundary for rejection is determined by finding the probability of sampling E independent uniform random variables whose empirical cdf has a larger test statistic than the set of p-values.

This last test will reject models that are either underfit or overfit, as an excess of overly small p-values or overly large p-values will lead to rejection. In general we only want to reject underfit models, but want to be aware if the proposed model appears to be overfit. To this end, a one-sided Kolmogorov-Smirnoff test can be used instead. In the one sided test the test statistic is the supremum of the raw difference, not the absolute difference, between the empirical cdf and the cdf of the uniform distribution. This test statistic is only large if there are too many small p-values. The model is rejected if the one-sided test statistic is implausibly large. An advantage of this test is that we can identify which edges lead to rejection by finding the edges with the smallest p-values. Another advantage of this test is that the infimum of the raw difference in the empirical cdf and the cdf of the uniform

distribution can be computed, and used as an indicator of overfitting. If the infimum is excessively large then the proposed model is an implausibly good fit to the data, so is likely overfit. As usual, what is “too large" is defined relative to the cdf of the test statistic when sampled from the uniform distribution.

In this chapter we will use the log-likelihood ratio as our test statistic in order to decide whether to accept or reject a model. However, the other test statistics will be computed to help evaluate the plausibility of the model given the data, whether it is over or under fit, and which edges are worst and best fit by the model.

The methods described above can be used to accept or reject a particular edge flow f. Our objective is to extend these tests to the composite hypotheses H_t and H_c. Both H_t and H_c assume that the true f lies in a subspace of $\mathbb{R}^E$. The transitive hypothesis H_t assumes that $f \in \text{range}(G)$. The cyclic hypothesis assumes that $f \in \text{null}(G^\intercal)$. To test H_t and H_c we apply the log-likelihood test to the maximum likelihood estimate f constrained to the appropriate subspace [193].

To find the maximum likelihood estimate for f constrained to a subspace we use a numerical optimizer to minimize the negative log-likelihood. The negative log-likelihood is a convex function, so minimizing the log-likelihood over a subspace is a convex optimization problem. Note that the constrained MAP estimator is given by minimizing the same cost function with γ added to the win record and 2γ to the event count on each edge. Let S be a subspace. Then a good initial guess for the MLE estimate of f constrained by the subspace is given by minimizing a quadratic approximation to the negative log-likelihood. The quadratic approximation to the negative log-likelihood about f_{MLE} is:

$$-\ln(\mathcal{L}(f|w,n)) \simeq -\ln(\mathcal{L}(f_{\text{MLE}}|w,n)) + \frac{1}{2}\sum_{k=1}^{E}\frac{w_k(n_k-w_k)}{n_k}(f-f_{\text{MLE}})_k^2 \qquad (5.7)$$

Thus an initial guess at the MLE edge-flow constrained to a subspace S can be given by solving a least-squares problem on the subspace. For example, suppose S is the perfectly transitive subspace, range(G). Then an initial guess at the MLE edge-flow constrained to range(G) is given by solving for the ratings r which minimizes $\sum_{k=1}^{E} \frac{w_k(n_k-w_k)}{n_k}((r_{j(k)} - r_{i(k)}) - \text{logit}(w_k/n_k))^2$, then setting $f = Gr$. Note that this approximation to the MLE edge flow is equivalent to a log-least squares rating where the the squared discrepancy on each edge is weighted by $w_k(n_k - w_k)/n_k$ instead of the standard weighting, n_k. Least squares ratings are widely used, for examples see [96, 128, 129, 126, 127, 123, 130, 131, 132]. Therefore, for the right choice of weights and edge-flow, standard least squares approaches can be considered approximations to the MLE ratings.

It follows that, when f_{MLE} is close to the perfectly transitive subspace then the quadratic approximation to the MLE ratings only differ from the Hodge rating (see equation 5.3) by the weighting on each edge. The approximate MLE ratings weight each edge by the asymptotic approximation to one over the variance in the posterior when using a uniform prior. Therefore the quadratic function minimized is equivalent to the Wald test statistic [187]. Under the quadratic approximation to the log-likelihood, the test statistic λ converges to the Wald test statistic, and minimizing negative log-likelihood amounts to minimizing the Wald statistic.

The Hodge ratings do not weight the discrepancy on each edge by the variance in the posterior on each edge since the Hodge rating system does not assume that the true tournament is perfectly transitive. Using the Hodge approach discrepancies in Gr and the edge-flow are assumed to come from the cyclic component, not sampling error, so are not re-weighted by a variance.

So, to find the MLE estimators for the edge-flow constrained to a subspace we first solve for $f_{\text{MLE}} = \text{logit}(w/n)$, then minimize the quadratic approximation to the log-likelihood

restricted to the subspace. This is a least squares problem so can be solved efficiently. Then the least squares approximant is used as an initial guess at the constrained MLE estimate and a numerical optimizer is used to find a solve for the constrained MLE estimate.

Once the constrained MLE estimate is found any of the test statistics can be evaluated on the constrained estimate. Note that the when using the AIC for a composite hypothesis we subtract $E - |S|$ from λ where $|S|$ is the dimension of the subspace S. Therefore, for the transitive hypothesis we subtract the dimension of the cycle space, $E - (V - 1)$, and for the cyclic hypothesis we subtract the dimension of the perfectly transitive subspace, V. Similarly, when comparing λ to its asymptotic χ^2 distribution, the χ^2 distribution should have degrees of freedom equal to to the dimension of the cycle space and transitive space respectively. Then the hypothesis $f \in S$ is rejected if the test statistic is too large, where the lower bound for rejection is fixed by the chosen significance level.

Note that this hypothesis testing framework is equivalent to the likelihood-ratio test widely used in logistic regression. In fact, if we let the predictor variable for a given competition event be $i(k)$ and $j(k)$, then the MLE ratings under the perfectly transitive hypothesis are equivalent to the regression coefficients in a logistic regression of the observed win record against the predictor.

Bounds on Measures

The hypothesis testing framework developed above can be extended to find bounds on the sizes of the components. Note that the primary advantage of the hypothesis testing approach is that a hypothesis is accepted if the maximum likelihood estimate constrained by the hypothesis could have credibly generated the observed data. Even if most credible models have a larger cyclic part than transitive part, if a model can be found with a small cyclic part that is accepted under the hypothesis test, then we cannot reject the hypothesis

that the cyclic part is at least that small. Thus, by using a hypothesis testing approach, biases associated with the differing dimension of the perfectly cyclic and transitive subspaces can be avoided. The Bayesian point and interval estimators depend on how most credible models behave, rather than if there is at least one credible model with a given behavior.

Let $H_{t\leq T}$, $H_{t=T}$ and $H_{t\geq T}$ correspond to the hypotheses that $||f_t||_2 \leq T$, $||f_t||_2 = T$ and $||f_t||_2 \geq T$ respectively. Let $H_{c\leq C}$, $H_{c=C}$ and $H_{c\geq C}$ correspond to the equivalent hypotheses on $||f_c||_2$. These hypotheses can be tested in almost exactly the same manner as described in the previous section. The regions defined by setting an upper bound on one of the components are convex regions, while the regions defined by putting a lower bound on the components are the union of two convex regions. Lastly, setting one of the measures equal to a fixed value specifes an affine subspace in $\mathbb{R}^E$, which is a convex set. Therefore solving for the maximum likelihood estimate constrained by putting a bound on the size of a component is a convex optimization problem.

Once the maximum likelihood estimate has been found, subject to the appropriate constraint, the hypothesis is either accepted or rejected by evaluating the p-value associated with the chosen test statistics. A bisection search can be used to find the value for the bound which separates hypotheses that are accepted and that are rejected.

For example, suppose we wanted to find the smallest C such that the hypothesis $H_{c\leq C}$ is accepted. Start by finding an upper $C = b$ such that the hypothesis is accepted, and a lower $C = a$ such that the hypothesis is rejected. Then set $C = (b - a)/2$, find the maximum likelihood estimate for f constrained to $||f_c||_2 \leq C$, evaluate the test-statistic on the estimate, and the p-value of that test statistic. If the p-value is above the desired significance accept the hypothesis and let $b = C$. Otherwise, reject the hypothesis and let $a = C$. This process can be accelerated at convergence by using a secant search instead of a bisection.

Using this technique we can find the smallest upper bound on the cyclic part that is credible, and the largest lower bound that is credible. Similarly we can find the smallest upper bound on the transitive part, and largest lower bound that is credible.

5.5 Elections

5.5.1 Motivation

Elections are an important example of competitive systems. In an election the votes cast, or opinions formed, by each voter may be interpreted as the outcome of a competitive event between the candidates or parties. Social choice theorists have repeatedly raised warnings about the possibility of voting paradoxes in which the “will of the majority" is ambiguous or ill-defined [194, 195, 98]. While some of these paradoxes may be of little importance to real elections, others have impacted observed election results (cf. [196]). For example, during the 2016 Republican primary repeated polls showed that Donald Trump held a plurality of the primary voters, but would have lost a head-to-head election against either Ted Cruz or Marco Rubio [197]. This is an example of the *Borda paradox* [98, 197], in which a candidate who would lose in a head-to-head election against some opponents beats those opponents with a plurality if the opponents split the same portion of the electorate. Concern about voting paradoxes is largely attributed to Arrow [194], who showed that no ranked voting system can simultaneously satisfy three fairness criteria. Arrow’s impossibility theorem is widely interpreted as showing that there is no universally fair method for picking a winner in an election, and is closely related to the Gibbard-Satterthwaite theorem [198], which demonstrates that there is no electoral system which is not susceptible to tactical voting [199, 200]. These paradoxes undermine the ideal

that election outcomes should reflect majority opinion [196]. It is important to note that Arrow's theorem does not imply that all election systems will fail to fairly pick a winner, only that for all election systems there exist some situations in which those systems fail. Election systems fail Arrow's fairness criteria when aggregate voter opinions involving the top candidates are cyclic [176]. This situation is an example of *Condorcet's paradox*.

Condorcet's paradox occurs when there is no *Condorcet winner* - a candidate who would defeat any other candidate in a head-to-head election [98]. If there is no Condorcet winner, then for any winner there is an opponent that a majority of the electorate prefers. This paradox arises when individual voter preferences lead to cyclic aggregate preferences among the leading candidates [98]. Consider an election among three candidates (A, B, and C) with three voters. If the first voter prefers A to B to C, the second prefers B to C to A, and the third C to A to B, then A would win a head-to-head election against B, B would win a head-to-head election against C, and C would win a head-to-head election against A [196]. Voter cycles have been observed in a number historical case studies. These include: voting in the House of Representatives and Senate on the annexation of Texas [168] and the subsequent status (free or slave) of land gained after the Mexican-American war [98, 169], voting on revenue bills in the House of Representatives in 1932 [167], voting in the Canadian parliament on abortion reform in 1988 [166], and public opinion on intervention in Kuwait preceding the Gulf War [165]. In each of the legislative examples the presence of a voting cycle meant that the legislative outcomes were largely dictated by the voting agenda, often yielding significant influence to those with the power to set the agenda [168], or resulting in a gridlock which preserved the status quo [166]. An extensive review of similar case studies is provided by [98].

Despite these examples the broader empirical relevance of Condorcet's paradox is controversial [196, 172, 170, 176]. The relevance of case-studies is contested as they often

rely on reconstructed voter preferences (cf. [169, 178]) that may be debated [201], and, at best, provide anecdotal evidence of the paradox [170]. Consequently, a number of authors have attempted to evaluate the frequency of voting cycles among large electorates using empirical preference data [196, 170, 175, 202, 176]. Combined, these studies indicate that Condorcet's paradox rarely occurs in large electorates, with most studies finding few if any cycles. Van Deemen [202] and Gehrlein [173] tabulate the outcome of every empirical search for Condorcet's paradox. They find that roughly ten percent of the studied elections (25 out of 265) exhibit the paradox [176], and that the paradox occurs more frequently in small electorates than large electorates.

This observation is in apparently stark contrast to axiomatic social choice theory, which emphasizes the "impossibility" of aggregating voter opinion. In particular, classical calculations and in-silico experiments indicate that either the more candidates, or the more voters, the higher the chance of cycles [203]. Since cycles are rarely observed in large elections the theory been criticized for overstating the prevalence of cycles [204, 171]. That said, there is extensive theory to explain why and when cycles are not expected to occur. Under certain domain restrictions on voter preferences it is possible to guarantee that aggregate voter preferences are transitive [195, 205]. For example, if the choices in an election can be arranged on a single axis, and all voter preferences are single peaked then the aggregate preferences are necessarily transitive and the Condorcet winner is the favorite of the median voter [195]. These domain restrictions are frequently violated in empirical studies so do not constitute a plausible explanation for the infrequency of cycles [172, 171].

An alternate body of theory considers the probability of observing cycles under specified assumptions about the distribution of voter preferences. These assumptions define a "culture". The classic observation that the probability of cycles increases with additional voters is based on assuming an *impartial culture* in which all voter preferences are

equally likely. In an impartial culture voter preferences are highly heterogeneous, which makes aggregation difficult. Under more realistic assumptions, voter preferences are more homogeneous and mutually correlated, which makes aggregation easier and reduces the chance of observing cycles [204]. Thus the primary distinction between existing theory and observed elections is mainly a matter of emphasis. Theory emphasizes the possibility of worst case scenarios. Empirical studies show that, while cycles can, and do, appear in influential elections (cf. [177]), they do so infrequently, especially in large electorates. The rarity of cycles in large elections does not necessarily violate theory, rather it reflects the distinction between in-silico voters and real voters. What is needed to advance the theory is an understanding of the distribution of voter preferences in real elections [204, 206], and a measure of heterogeneity in voter opinion which can be related to an expected prevalence of cycles.

The HHD is a natural analytic framework for answering the questions raised by this discussion. First, cycles play an essential role in social choice theory, and have been the subject of protracted debate, so a decomposition which isolates cyclic preferences could be useful. Second, most real elections are transitive, which begs the question, are most real elections transitive because aggregate voter opinion is described by a predictive rating system, or simply because any cyclic component of voter preferences is small relative to the transitive part? Demonstrating the former would indicate that voter preferences satisfy surprisingly strong assumptions. Demonstrating the latter would show that aggregate voter opinion is mostly described by a predictive rating, but variation in voter opinions leads to a latent cyclic component that is usually hidden by a larger transitive component. Detecting and characterizing the size of this component would quantify how close aggregate voter opinions are to perfect transitivity, and how large the intransitive inconsistencies in voter opinion are. The larger the cyclic component the more likely intransitivity, and, as a

consequence, voting paradoxes. Moreover, in Chapter 4 we demonstrated that, the expected size of the cyclic and transitive components are related to the correlation in the log-odds that competitor A beats B with the log-odds that competitor A beats C. This correlation is a measure of how consistent voter preferences are, so could be used as a metric for the homogeneity of opinion in an electorate. Therefore we apply the HHD to election data in order to: 1. identify any cyclic preferences in voter opinion, 2. quantify and compare the sizes of the cyclic and transitive parts, 3. relate the size of the cyclic component to statistical assumptions about the electorate, 4. test the hypothesis that aggregate voter opinion is perfectly transitive and any observed cyclic part is a result of sampling error.

5.5.2 Data

In this chapter we consider data drawn from three different election systems. The examples considered are: eight Danish parliamentary elections ranging from 1973 to 2005, four Dutch parliamentary elections ranging from 1982 to 1994, and 12 American presidential elections ranging from 1968 to 2020. The Danish examples are based on work by Kurrild-Klitgaard [196] and the Dutch examples are based on work by van Deemen [202]. The American examples are inspired by two studies [207, 208] which examined the 1968, 1972, 1976, 1980, and 1992 elections. We go beyond these examples to consider modern Presidential elections. All data on the American presidential elections was drawn from the American National Election Study (ANES). All of the data sets used involved over 1000 poll respondents.

A standard challenge when searching for cycles in voting preferences is a lack of appropriate polling. To test for Condorcet's paradox, or intransitivity in general, the investigator must be able to predict the outcome of a head-to-head election between any pair of candidates. This requires asking poll respondents to either rank or rate the candidates

[197]. This sort of polling is rarely performed [172, 196]. For example, in the 2016 Republican primary only five major published polls provided head-to-head comparisons between the candidates, and only two provided head-to-head comparisons between more than the leading candidate and runner-ups [197]. While rare, workable polling data exists for certain elections. A number of national election surveys ask poll respondents to rate candidates, or parties, and those ratings can be used to estimate the preference order of each poll respondent (cf. [196]). For example, since 1968 the ANES has asked poll respondents to rate important political figures from 0 to 100 on a “feeling thermometer". ANES data is easily accessible and has been used in other social choice studies [207, 208, 204]. We assume that if a respondent rated candidate A over B (with a sufficient margin) that the respondent preferred A to B. This technique is widely used in social choice literature, particularly in empirical studies of large electorates.

In our study of Dutch and Danish elections we use the published preference matrices in [202] and [196]. Depending on the year these elections involve nine to thirteen parties, and average over a thousand responses per pair. Danish and Dutch parliamentary elections are particularly interesting election systems as both involve many parties, which fracture the electorate, thereby introducing more possibilities for inconsistency in aggregated voter preferences. In our study of American elections we use the published preference matrices provided by [207] and [208] (1968, 1972, 1976, 1980, 1984, 1992), and use ANES data to form preference matrices for elections in 1988, 1996, 2000, 2008, 2016, and 2019. We also use the ANES data to validate our procedure for building preference matrices from thermometer polling against the preference matrices published in [207] and [208]. These preference matrices typically involve the two major party candidates, any major third party candidates, and important challengers to either of the major party candidates. The ANES data includes polls performed both before and after the election, so for elections studied

using ANES data we consider voter opinion before and after the election. The ANES data also includes pilot studies run before the primaries are complete. We use data from the ANES pilots in 2016 and 2019 to study competition among the primary candidates in the Republican party in 2016, and Democratic party in 2019. Combining these polls can give different snapshots of the a single election. For example, in 2016 we consider the public opinion regarding the major candidates in the Republican party, public opinion about Donald Trump, Hillary Clinton, and Bernie Sanders preceding the election, and public opinion regarding Donald Trump, Hillary Clinton, Jill Stein, and Gary Johnson before and after the election. As in the European examples these data sets contain, on average, over a thousand responses per pair of candidates.

In order to treat the poll results using the HHD we need to define what is meant by a competition event. Here we define a competition event between candidates A and B to be the opinion of a single, randomly chosen, voter regarding A and B, and say that A beats B if the randomly chosen voter prefers A to B. We assume that, for the population sampled, there is some probability that a randomly chosen voter will prefer A to B, so that the actual number of voters sampled who preferred A to B is a realization of a binomial random variable with unknown win probability. An alternative approach would be to consider the election as a competition event, rather than the opinion of a randomly chosen individual, however this would require modelling head-to-head elections which do not take place in most of the systems considered. Election rules are often idiosyncratic, as in American presidential elections, so modelling competition at the level of the election may not accurately reflect the underlying opinions of the electorate, and would require methods tailored to each system separately.

5.5.3 Results

Transitivity and Hypothesis Testing

All examples studied were transitive. That is, for every example there existed a ranking of the candidates such that, if A ranked higher than B, then more poll respondents preferred A to B in a head-to-head comparison. Therefore Condorcet's paradox was not observed in the aggregate voter preferences in any of the 24 elections considered. Moreover, the posterior probability that each example was drawn from a set of underlying win probabilities that were transitive was greater than 0.775 (2019 pilot study) for all elections, and averaged 0.94, 0.94, and 0.98 for the Danish, Dutch, and American elections respectively. This motivates testing the stronger hypothesis that aggregate voter opinion was not only transitive, but perfectly transitive, in each of the 24 elections. If we can reject the hypothesis that aggregate voter opinion is perfectly transitive then there exists latent cyclic structure in the electorate's preferences. The larger this component the less accurately the electorate's preferences can be predicted by assigning a rating (popularity) to each candidate/party. This reflects heterogeneity in preference across individuals, thus the potential for voting paradoxes in future elections.

The perfectly transitive hypothesis H_t is rejected with high confidence in all 4 Dutch elections and all 8 Danish elections. The perfectly transitive hypothesis is rejected in approximately half of the American elections, often with much less confidence than in the Dutch or Danish elections. A summary of the test statistics is provided in Table 5.1. Note that for all Danish and Dutch elections the difference in AIC between the MLE model and the MLE model under the perfectly transitive hypothesis is positive and large (order 10^2). This indicates that the MLE model under the transitive hypothesis is much less likely than the MLE model, and that it is unlikely that the difference in likelihood is accounted for by

Danish	1973	1975	1977	1979	1994	1998	2001	2005
log(Likelihood Ratio)	331.9	1111.0	735.3	1392.6	442.0	985.0	677.5	1065.6
d.o.f.	45	45	45	45	28	45	21	36
AIC	241.9	1021.0	645.3	1302.6	386.0	895.0	635.5	993.6
Sample p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Dutch	1982	1986	1989	1994
log(Likelihood Ratio)	533.5	543.4	230.0	395.7
d.o.f.	66	55	28	28
AIC	401.5	433.4	174.0	339.7
Sample p-value	0.000	0.000	0.000	0.000

American (post)	1968	1972	1976	1980	1984	1988	1992	1996	2000	2008	2016
log(Likelihood Ratio)	0.6	13.7	9.7	21.0	7.2	1.4	2.8	28.4	6.2	8.6	3.4
d.o.f.	1	1	1	6	1	1	1	3	1	1	3
AIC	-1.4	11.7	7.7	9.0	5.2	-0.6	0.8	22.4	4.2	6.6	-2.6
Sample p-value	0.846	0.003	0.022	0.015	0.067	0.692	0.419	0.000	0.101	0.034	0.746

Table 5.1: Test statistics for the perfectly transitive hypothesis. The log likelihood ratio is (twice) the difference in the log likelihood of the MLE estimate for the edge flow given the data and the MLE estimate for the edge flow constrained to the perfectly transitive (arbitrage free subspace), range$\{G\}$. A large log likelihood difference indicates that the MLE estimate under the perfectly transitive hypothesis is much less likely than the MLE estimate when the edge flow is unconstrained. The perfectly transitive subspace only has $V - 1$ degrees of freedom, while the unconstrained edge flow has E degrees of freedom. The difference in degrees of freedom is provided in the row labelled d.o.f. (degrees of freedom). The difference in degrees of freedom is the dimension of the loop space. In general, the larger the difference in d.o.f. the larger the log likelihood difference (asymptotically the log-likelihood ratio is χ^2 distributed with parameter equal to the d.o.f. difference if the perfectly transitive hypothesis is true). The difference in the Aikake Information Criterion (AIC) between the MLE model and the MLE model constrained to the perfectly transitive response is provided in the next row. This equals the log likelihood ratio minus twice the d.o.f. If the AIC difference is negative then the MLE model is likely overfit, and the difference in log-likelihood ratio can be accounted for by the difference in d.o.f. A large (positive) difference in AIC indicates that the difference in likelihood between the MLE model and MLE model under the hypothesis cannot be accounted for by the difference in d.o.f. The sample p-value (probability that, if the MLE model under the hypothesis, a win record more or equally unlikely than the observed data would be sampled) is provided in the last row. If the sample p-value is greater than or equal to significance 0.05 then the hypothesis is accepted. Otherwise the hypothesis is rejected. All of the values for the American elections use post-election polls. Test statistics for pre election polling, and pilot studies are provided in the supplement.

overfitting of the unconstrained MLE model. The sample p-values (probability that, if the MLE estimate win probabilities given the perfectly transitive hypothesis were the true win probabilities, we would sample a win record more or equally as unlikely as the actual win record) for all of the Danish and Dutch elections are less than 10^{-3}. In contrast, for most of the American elections the AIC difference is small, and occasionally is negative, indicating that the unconstrained MLE model might be overfit. Similarly, the sample p-values are relatively large, indicating that the perfectly transitive hypothesis is more plausible for the American elections. The hypothesis is accepted in 1968, 1984, 1988, 1992, 2000, and 2016, and is rejected in 1972, 1976, 1980, 1996, and 2008. Notice that the years with a large sample p-value match the years with a small AIC difference and small log likelihood ratio. Also note that in 1984 and 2000 the hypothesis is accepted with a marginal sample p-value, and in 1976, 1980, and 2008 the hypothesis is rejected with relatively large sample p-values. The results provided in the table are only for post election polls.

Therefore, the Danish and Dutch examples clearly are not perfectly transitive, while American elections are close to perfectly transitive, and may be perfectly transitive in certain years. It is possible that American elections are close to, but not perfectly transitive, and the sample size of the polls was not large enough to reject the hypothesis. Alternatively it is possible that voter preferences were perfectly transitive in some years, and not others.

Pre-election polling results are not reported here, but have similarly mixed results, often matching the results in the corresponding post election poll. In 1980 the pre-election poll results are noticeably different than the post election poll results. Early in the election Carter was favored over Reagan by a small margin [207], and the perfectly transitive hypothesis is accepted with sample p-value 0.35. For the 1968, 1980, and 1992 election Abramson [207] provides additional preference matrices that include preferences of only respondents who planned on voting, and validated voters. When controlling for voting

intention the test statistics change (pre-election 1980 sample p-values are 0.35, 0.28, 0.19 for all respondents, respondents who planned on voting, and voters respectively) but not enough to change the conclusions of the hypothesis test. Pilot study and pre-election data was also considered for the 2000, 2016, and 2020 elections. The 2000 pre-election survey considered Al Gore, George Bush, Ralph Nader, John McCain, and Bill Bradley, the 2016 pilot considered Jeb Bush, Donald Trump, Ben Carson, Marco Rubio, Ted Cruz, Carly Fiorina, Hillary Clinton, and Bernie Sanders, and the 2019 pilot considered Donald Trump, Joe Biden, Bernie Sanders, Elizabeth Warren, Pete Buttigieg, and Kamala Harris. We reject the perfectly transitive hypothesis with sample p-value 0.004 for the 2000 pre-election survey. We accept the perfectly transitive hypothesis with sample p-value 0.876 for the 2016 Republican primary, but reject the perfectly transitive hypothesis with sample p-value less than 10^{-3} for the 2016 pilot if Sanders and Clinton are considered. Similarly, we accept the perfectly transitive hypothesis for the 2019 Democratic primary with sample p-value 0.577, but reject the perfectly transitive hypothesis with sample p-value less than 10^{-3} if Donald Trump is included in the preference matrices. So, despite the relatively large number of candidates in the 2016 Republican primary, and 2020 Democratic primary, aggregate voter opinion between the primary candidates was plausibly perfectly transitive, but voter opinion across parties is not perfectly transitive. This indicates that what makes a candidate popular within a party differs from what makes a candidate popular when competing against candidates from other parties, perhaps because voter opinion within a party is more homogeneous than voter opinion across parties, or because voter preference regarding similar candidates depend on different criteria than voter preference regarding disparate candidates.

Taken together these results suggest that voter preferences are close to perfectly transitive in most American presidential elections, are closer in primaries, and that how close

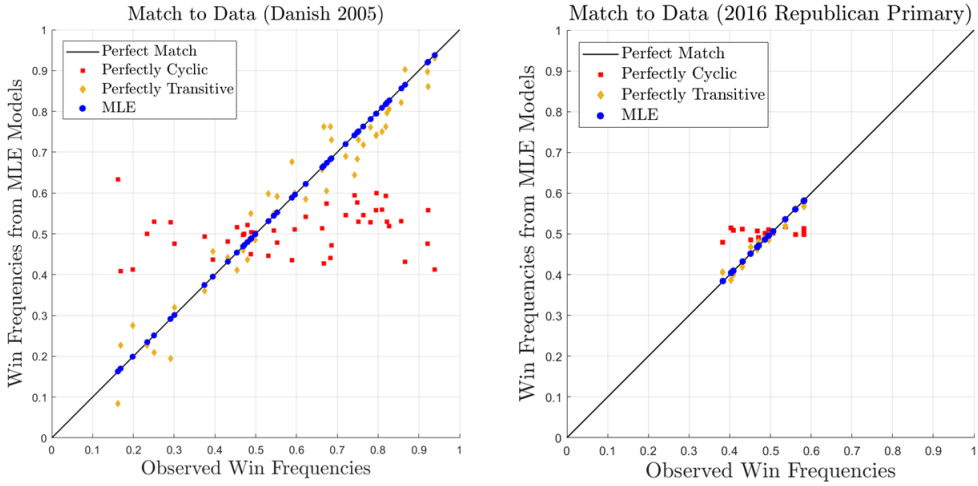

Figure 5.3: Comparison of observed win frequencies per edge (fraction of poll respondents who prefer A to B) against predicted win frequencies under the MLE model (blue circles), MLE model satisfying H_t (gold diamonds), and MLE model satisfying H_c (red squares). The left panel represents the 2005 Danish election and the right panel represents the 2016 Republican primary. The left panel is representative of the 12 Danish and Dutch elections, while the right panel is representative of the American elections which are plausibly perfectly transitive. Each scatter point represents an individual edge (pair of competitors), and the vertical distance of the scatter point from the black diagonal is the discrepancy between the predicted win frequency (under the model) to the observed win frequency.

voter preference is to perfectly transitive depends principally on which candidates are considered, not whether the poll respondents are restricted by voting behavior. They also indicate that the plausibility of the perfectly transitive hypothesis does not change significantly between pre and post election surveys except in elections for which a candidate's popularity changed dramatically over the course of the election (Carter and Reagan in 1980).

In contrast, for all of the elections studied we reject the hypothesis that the election is perfectly cyclic with high confidence (sample p-value less than 10^{-3}). In every election considered the perfectly cyclic hypothesis has notably large test statistics, and a small sample p-value. Figure 5.3 compares the observed win frequencies on each edge (fraction

of poll respondents who prefer A to B) to the predicted win frequencies using the MLE model, MLE model under the perfectly transitive hypothesis, and MLE model under the perfectly cyclic hypothesis. The left panel is for the 2005 Danish parliamentary election, and the right panel is for the 2016 Republican primary. These two examples are chosen as the left panel is representative of the Danish and Dutch elections, while the right panel is representative of the American elections that are plausibly perfectly transitive. Blue circles are the MLE model, gold diamonds represent the MLE model under the perfectly transitive hypothesis, and red squares represent the MLE model under the perfectly cyclic hypothesis. Note the marked disparity between the observed win frequencies and the win frequencies predicted using the MLE model with the perfectly cyclic hypothesis. Note that in both panels the MLE model under the perfectly transitive hypothesis gives a good, but not perfect match to the data. The distance from the black diagonal is the discrepancy between the data and the model. Note that this discrepancy is much more pronounced in the Danish election example than in the Republican primary. In the former case the discrepancy is too large to be plausibly explained by sampling error, while in the latter case the discrepancies are small, and could be the result of sampling error.

The fact that all of the elections are transitive, yet few are plausibly perfectly transitive, and in all cases the perfectly transitive hypothesis outperforms the perfectly cyclic hypothesis suggests that the elections likely have a small, but nonzero, cyclic component. The cyclic component accounts for the discrepancies between the predicted “win frequencies" (fraction of respondents who preferred A to B) given the MLE model under the perfectly transitive hypothesis and the observed win frequencies that cannot be plausibly explained as sampling error. Since we expect the cyclic component to account for the discrepancies between the observed win frequencies, and the predicted win frequencies under the perfectly transitive hypothesis, we identify the edges (pairs of competitors) for which

the observed preferences are the least likely given the preferences predicted by the MLE estimate under the perfectly transitive hypothesis. For each edge we compute a two-sided p-value (probability on the given edge of sampling preferences more or equally unlikely as the actual preferences had the MLE model under the perfectly transitive hypothesis been the truth), and rank the edges in increasing order. Next, to estimate the cyclic component and transitive component of the aggregate preferences we apply the HHD to the estimated log-odds that a randomly sampled voter would have preferred candidate A to candidate B across all pairs of candidates. As an example, Figure 5.4 shows the predicted win frequency and observed win frequency for the ten pairs of parties with the smallest p-values for the Danish parliamentary election of 1973.

The HHD: Ratings

To estimate the cyclic component and transitive component of the aggregate preferences, we apply the HHD to the estimated log-odds that a randomly sampled voter would have preferred candidate A to candidate B across all pairs of candidates. This produces a rating and ranking of the candidates/parties in the election, and a pair of edge flows, one which is perfectly transitive and one which is perfectly cyclic.

In general, the rankings produced by applying the HHD to the estimated edge flow matched the outcomes of the elections. Out of all 28 preference matrices considered for the 12 American Presidential elections the candidate ranked first using the estimated ratings matched the winner of the election in all but 6 cases (1980 pre-election, 2000 pre-election, and all four preference matrices considered for 2016). Moreover, in all but 4 cases (2008, all three preference matrices considered in 2016 that involve the general election) the top two candidates were the Republican and Democratic nominees. When the HHD is applied to the pre-election 1980 preference data (not accounting for voting intention) Carter is ranked

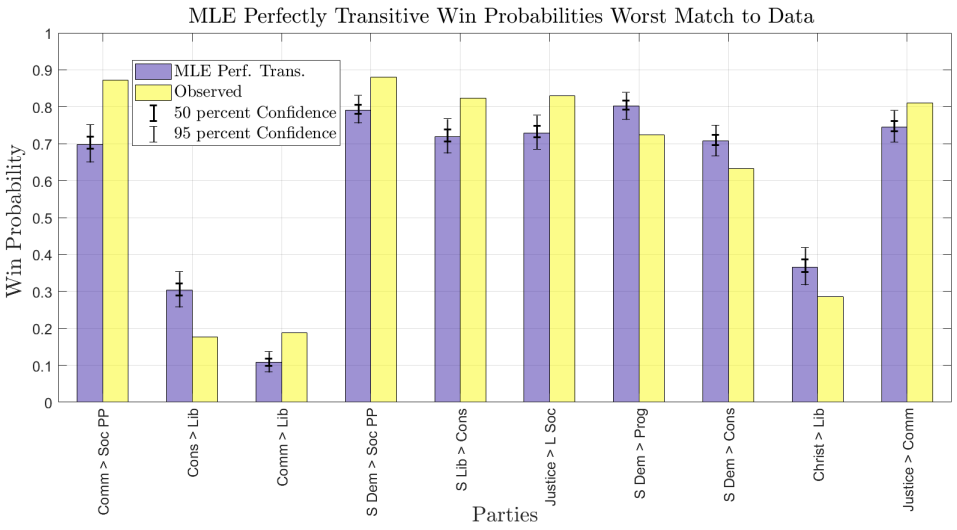

Figure 5.4: Comparison of observed win frequencies (yellow bars) against predicted win frequencies under MLE model satisfying H_t (blue bars) for the ten edges with the largest discrepancy between prediction and observation. The error bars represent the range of observed win frequencies that could have plausibly been observed had the blue bars represented the actual win probabilities. The inner pair of error bars represent the 50 percent confidence interval, and the outer pair represent the 95 percent confidence interval, on the sampled win frequency had the perfectly transitive model been the truth. Notice that the observed win frequencies fall outside the 95 percent confidence interval on all ten edges.

ahead of Reagan. This matches the trajectory of the election as described by [207]. Carter held a slight polling advantage at the start of the election, but rapidly lost popularity over the course of the election, culminating in his defeat in November. In the 2000 pre-election survey John McCain is ranked first, followed by Al Gore, and then George H.W. Bush. John McCain won the second most delegates in the 2000 primary, and ran as a moderate candidate. Thus it is not be surprising that he was preferred by a set of poll respondents who included both Republicans and Democrats. The 2000 post-election survey ranks Bush ahead of Gore ahead of Nader. Moreover, the estimated ratings using both the pre and post election surveys are remarkably close. Before the election the estimated ratings for Bush and Gore are 0.003 ± 0.01 and 0.068 ± 0.01, and after the election the estimated ratings are

0.076 ± 0.01 and 0.066 ± 0.01. This leads to considerable uncertainty in the posterior distribution for the rank ordering of Bush and Gore. Once McCain is removed from consideration the Spearman and Kendall rank correlation coefficients for the 2000 election are only 0.70 and 0.78, the smallest seen across all 28 elections considered excluding 2016. This uncertainty reflects the historically close outcome of the 2000 election.

The 2016 election is particularly notable in how starkly the estimated rankings differ from the election results. Remarkably, Donald Trump is the lowest rated candidate using the estimated edge flow no matter the data source (pilot, pre, or post election data), and no matter the combination of candidates considered (Republican primary candidates, Republican primary candidates plus Hillary Clinton and Bernie Sanders, or general election candidates Hillary Clinton, Donald Trump, Gary Johnson, and Jill Stein). In all cases Trump is the Condorcet loser - a candidate who would lose a majority head-to-head election against any other candidate - in both the Republican primary, and the general election surveys. This confirms polling results reported in [197], in which it was argued that Trump would have lost a head-to-head election against Marco Rubio or Ted Cruz, despite winning a plurality of the vote. As for McCain, it should be noted that the polling respondents include members of both parties, so the primary results reflect national opinions of the Republican candidates rather than Republican opinion. Therefore it is possible that Trump is ranked last among the Republican primary candidates due to his unpopularity among Democrats. It should also be noted that the Spearman and Kendall rank correlations are very small in this case (0.27 and 0.22 respectively), indicating that there is considerable uncertainty in this rank order. Notably the Condorcet winner, and highest ranked candidate, when considering all candidates included in the pilot study is Bernie Sanders, followed by Marco Rubio. Hillary Clinton is ranked fourth in the pilot study. The pre and post election surveys only considered the two major party nominees, along with major third

party candidates Jill Stein and Gary Johnson. In both pre and post election surveys Gary Johnson is the Condorcet winner and highest ranked candidate, followed by Jill Stein, then Hillary Clinton, then Donald Trump. Of all 12 American elections considered 2016 is the only year in which a third party candidate is ranked first, and is the only year in which third party candidates are not the lowest ranked of all candidates (Wallace, Anderson, Perot and Nader all clearly rank behind the major party candidates). The only other election in which the Democratic and Republican nominees are not ranked first and second is 2008, because Hillary Clinton is ranked ahead of John McCain. In both pre and post election surveys the Kendall and Spearman rank correlations are relatively large (0.92, 0.94, 0.95, and 0.97) so the reversal in rank order cannot be attributed to uncertainty in the posterior distribution of rank orderings. This highlights the historic unpopularity [209] of both major party candidates in 2016.

Therefore, while the estimated rank orders using the HHD do not match the election outcomes in all 12 elections, the elections in which the rank ordering does not match the election outcomes are elections in which some of the candidates' popularity changed significantly over the course of the election (1980), were historically close (2000), or for which both major party candidates were historically unpopular (2016).

A more detailed comparison can be made for the Danish elections, since seats in the Danish parliament are distributed according to a proportional representation (PR) system. The parties can then be ranked according to the number of seats they win in the election, and the rank order using PR can be compared to the ranking predicted by the HHD. Moreover, since all eight Danish elections were transitive, the parties can be ranked unambiguously using pairwise majority relations (MR). The ranked parties and ratings are provided in Table 5.2. Orderings according to MR and PR are reproduced from [196]. Due to the large sample size the standard deviation in the ratings is small ± 0.01, and, as a result, the

rankings can be predicted confidently (average Kendall and Spearman rank correlations of 0.92 and 0.93 respectively). Note that the rankings associated with the HHD closely match the rankings given by MR. MR is equivalent to the HHD if the network is perfectly transitive, or all of the win probabilities are rounded to zero or one. Discrepancies in the rank order between MR and HHD are accounted for by the cyclic component of the HHD. Note that the range of ratings assigned to the most and least popular parties remained fairly consistent across the elections, with occasional outlying minority and majority parties. This consistency indicates that the distribution of popularity among parties remained reasonably consistent across the eight elections.

Even though each election was transitive, the outcome using PR was different from the outcome using MR in all elections [196]. A few minority parties consistently performed better using PR than they would have using MR or as predicted using the Hodge ratings. In particular, the Socialist People's Party, Communist party, and the Socialist Unity List party frequently won more seats than parties they would have lost to in a head-to-head election [196]. In 1973, 1975, and 1977 the Communist party would have lost a head-to-head election against any other party in the election, but won more seats than 2, 3, and 4, other parties respectively. Even more strikingly, in 1973 the Progress party was ranked 9th, and 8th using the Hodge ratings, or MR, but won the second most seats, displacing 6 other parties. The Justice party, on the other hand, consistently underperformed. In 1975 the Justice party would have ranked 6th out of 11 parties using the Hodge ratings, or 7th out of 11 using MR, but came in last in the election winning zero seats and losing to 4 parties it would have beat in a head-to-head election [196]. The general success of 'fringe' parties in Denmark's elections [196] (see the Socialist People's Party, Communist Party, Socialist Unity List party, Progress Party, and Danish People's party in Table 5.2) leaves open the possibility that there is a considerable cyclic component in the Danish electorate's

	1973				1975				1977			
Rank	HHD	Rating	MR	PR	HHD	Rating	MR	PR	HHD	Rating	MR	PR
1	S Dem	0.50	S Dem	S Dem (46)	Lib	0.49	Lib	S Dem (53)	S Dem	0.79	S Dem	S Dem (65)
2	S Lib	0.49	S Lib	Prog (28)	S Dem	0.47	S Dem	Lib (42)	C Dem	0.25	C Dem	Prog (26)
3	Lib	0.43	Lib	Lib (22)	Christ	0.36	Christ	Prog (24)	Cons	0.20	S Lib	Lib (21)
4	Christ	0.16	Christ	S Lib (20)	S Lib	0.31	S Lib	S Lib (13)	S Lib	0.18	Cons	Cons (15)
5	C Dem	0.13	C Dem	Cons (16)	Cons	0.11	Cons	Cons (10)	Lib	0.05	Justice	C Dem (11)
6	Cons	0.00	Cons	C Dem (14)	Prog	-0.02	Prog	Soc PP (9)	Justice	0.00	Lib	Soc PP (7)
7	Justice	-0.07	Justice	Soc PP (11)	Justice	-0.1	C Dem	Christ (9)	Christ	-0.06	Christ	Comm (7)
8	Soc PP	-0.18	Prog	Christ (7)	C Dem	-0.17	Justice	Comm (7)	Soc PP	-0.1	Soc PP	S Lib (6)
9	Prog	-0.23	Soc PP	Comm (6)	Soc PP	-0.2	Soc PP	C Dem (4)	Prog	-0.36	L Soc	Justice (6)
10	L Soc	-0.59	L Soc	Justice (5)	L Soc	-0.56	L Soc	L Soc (4)	L Soc	-0.44	Comm	Christ (6)
11	Comm	-0.64	Comm	L Soc (0)	Comm	-0.68	Comm	Justice (0)	Comm	-0.53	Prog	L Soc (5)

	1979				1994				1998			
Rank	HHD	Rating	MR	PR	HHD	Rating	MR	PR	HHD	Rating	MR	PR
1	S Dem	0.71	S Dem	S Dem (68)	S Dem	0.49	S Dem	S Dem (62)	S Dem	0.58	S Dem	S Dem (63)
2	S Lib	0.42	S Lib	Cons (22)	Cons	0.34	Lib	Lib (42)	Lib	0.46	Lib	Lib (42)
3	Lib	0.39	Lib	Lib (22)	Lib	0.31	Cons	Cons (27)	Cons	0.33	Cons	Cons (16)
4	Cons	0.24	Cons	Prog (20)	S Lib	0.18	S Lib	Soc PP (13)	C Dem	0.27	C Dem	Soc PP (13)
5	Soc PP	0.08	Soc PP	Soc PP (11)	C Dem	0.02	C Dem	Prog (11)	S Lib	0.21	S Lib	DPP (13)
6	Justice	-0.06	Justice	S Lib (10)	Soc PP	-0.02	Soc PP	S Lib (8)	Soc PP	0.21	Soc PP	C Dem (8)
7	Christ	-0.18	Christ	C Dem (6)	Christ	-0.24	Christ	Soc UL (6)	Christ	-0.01	Christ	S Lib (7)
8	C Dem	-0.19	C Dem	L Soc (6)	Prog	-0.40	Prog	C Dem (5)	DPP	-0.3	DPP	Soc UL (5)
9	L Soc	-0.26	L Soc	Justice (5)	Soc UL	-0.69	Soc UL	Christ (0)	Soc UL	-0.39	Prog	Christ (4)
10	Prog	-0.53	Comm	Christ (5)					Prog	-0.4	Soc UL	Prog (4)
11	Comm	-0.63	Prog	Comm (0)					Dem R	-0.94	Dem R	Dem R (0)

	2001				2005			
Rank	HHD	Rating	MR	PR	HHD	Rating	MR	PR
1	Lib	0.39	Lib	Lib (56)	S Dem	0.51	Lib	Lib (52)
2	S Dem	0.37	S Dem	S Dem (52)	Lib	0.37	S Dem	S Dem (47)
3	Cons	0.17	Cons	DPP (22)	Cons	0.3	Cons	DPP (24)
4	S Lib	0.05	S Lib	Cons (16)	S Lib	0.3	S Lib	Cons (18)
5	Soc PP	0.01	Soc PP	Soc PP (12)	Soc PP	0.13	Soc PP	S Lib (17)
6	Christ	-0.08	Christ	S lib (9)	DPP	-0.11	DPP	Soc PP (11)
7	DPP	-0.34	DPP	Soc UL (4)	Soc UL	-0.19	Soc UL	Soc UL (6)
8	Soc UL	-0.58	Soc UL	Christ (4)	Christ	-0.22	Christ	Christ (0)
9					C Dem	-0.28	C Dem	C Dem (0)
10					Minority	-0.82	Minority	Minority (0)

Table 5.2: Rating and ranking of Danish political parties. Each set of four columns corresponds to a year. The columns labelled HHD list the parties according to the Hodge ranking of the estimated edge flow f_{exp}. The corresponding ratings are provided in the neighboring columns (standard deviation 0.02 in 1973, and 0.01 in all subsequent elections). Next the parties are ranked according to pairwise majority relations (MR). Lastly the parties are listed by the number of seats won in the election via proportional representation (PR).

aggregate preferences. Since each election was transitive we expect this component to be smaller than the transitive component. We also expect the cyclic component to be largest on edges associated with minority parties.

The HHD: Measures and Characterization

For each election we estimated the sizes of the transitive and cyclic components. First, the sizes of the components in the estimated flow were computed. Then the posterior distribution for each measure was approximated by sampling, and a credible interval for the size of each component was estimated. The standard deviation in the posterior was also computed. To conclude, the largest lower bound, and smallest upper bound, on the size of each component such that the MLE estimate satisfying the bound passed the hypothesis test were computed. The sizes of the transitive and cyclic component for all eight Danish elections are reported in Table 5.3.

Despite changes in which parties participated in the elections, the number of parties in the elections, and over thirty years between the first and last election studied, the sizes of the transitive and cyclic components remain remarkably consistent across all eight elections. The transitive component remained between 0.53 and 0.57 with only two exceptions, and the cyclic component remained between 0.14 and 0.18 in all eight years. The estimated proportion of competition that is cyclic remained consistent across all eight Danish elections. The estimated relative sizes of the cyclic component are 0.27, 0.31, 0.25, 0.29, 0.27, 0.22, 0.30, and 0.24, all with standard deviation 0.01. This suggests that Danish elections, at least from 1970 - 2005, are characterized by a latent cyclic component that is about a quarter the size of the transitive component.

Like the Danish elections, all four Dutch elections have transitive and cyclic components with consistent sizes, and the standard deviation in the posterior is relatively small.

Danish	Transitive			Cyclic		
Year	Estimate (± 0.01)	Credible Interval	Credible Bounds	Estimate (± 0.01)	Credible Interval	Credible Bounds
1973	0.56	$[0.55, 0.58]$	$[0.49, 0.65]$	0.15	$[0.15, 0.18]$	$[0.08, 0.24]$
1975	0.56	$[0.55, 0.57]$	$[0.51, 0.61]$	0.18	$[0.17, 0.19]$	$[0.13, 0.23]$
1977	0.53	$[0.52, 0.54]$	$[0.48, 0.58]$	0.14	$[0.13, 0.15]$	$[0.09, 0.19]$
1979	0.59	$[0.58, 0.59]$	$[0.55, 0.63]$	0.17	$[0.17, 0.19]$	$[0.13, 0.22]$
1994	0.54	$[0.53, 0.56]$	$[0.49, 0.60]$	0.15	$[0.14, 0.17]$	$[0.10, 0.21]$
1998	0.65	$[0.64, 0.66]$	$[0.60, 0.71]$	0.14	$[0.14, 0.16]$	$[0.10, 0.20]$
2001	0.47	$[0.46, 0.48]$	$[0.43, 0.51]$	0.15	$[0.14, 0.16]$	$[0.11, 0.19]$
2005	0.57	$[0.56, 0.58]$	$[0.52, 0.61]$	0.15	$[0.14, 0.15]$	$[0.11, 0.18]$
Dutch	Transitive			Cyclic		
Year	Estimate (± 0.01)	Credible Interval	Credible Bounds	Estimate (± 0.01)	Credible Interval	Credible Bounds
1982	0.84	$[0.82, 0.85]$	$[0.76, 0.94]$	0.15	$[0.15, 0.17]$	$[0.07, 0.25]$
1986	0.91	$[0.89, 0.93]$	$[0.81, 1.04]$	0.15	$[0.15, 0.18]$	$[0.08, 0.28]$
1989	0.86	$[0.85, 0.88]$	$[0.79, 0.94]$	0.12	$[0.12, 0.15]$	$[0.06, 0.20]$
1994	0.85	$[0.84, 0.87]$	$[0.79, 0.93]$	0.16	$[0.15, 0.19]$	$[0.10, 0.24]$

Table 5.3: Estimated sizes of the transitive and cyclic components of the log-odds edge flow (normalized by $\sqrt{E}$) for eight Danish elections and four Dutch elections. The standard deviation in the posterior for each component were 0.01 or less for all entries. The credible interval is the 95 percent credible interval approximated by sampling from the posterior. The credible bounds are the smallest upper bound, and largest lower bound, for which the hypothesis that the log-odds satisfy the constraint is accepted using the log-likelihood test with significance 0.05. This interval is always wider than the credible interval as it only requires that the MLE model satisfying the bounds is credible, not that the region inside the bounds contains most of the mass of the posterior distribution.

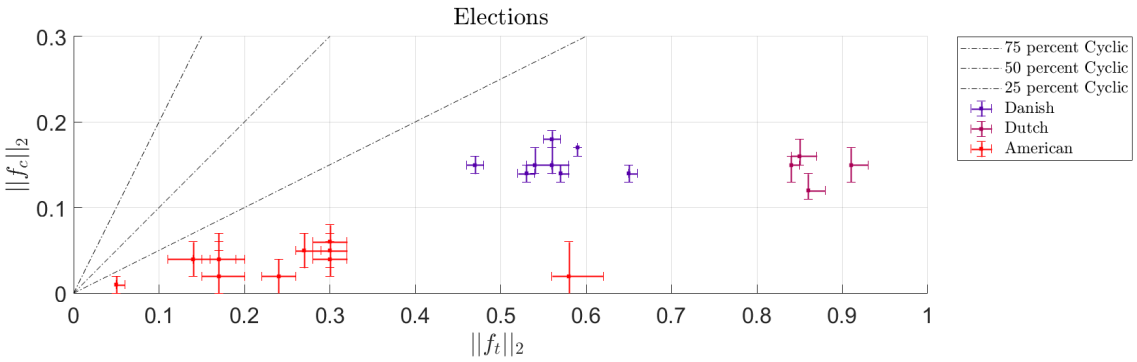

Figure 5.5: Estimated transitive and cyclic components normalized by $\sqrt{E}$ for the eight Danish elections (blue diamonds), four Dutch elections (purple circles), and twelve American elections (red squares). Error bars denote the credible interval for each measure. Note that the American elections (with the exception of 1968) cluster in the bottom left hand corner near the origin, the Danish elections cluster in the center of the figure, and the Dutch elections on the right.

Note that the cyclic component in the Dutch elections is the same size as the cyclic component in the Danish elections, but the transitive component in the Dutch elections is consistently larger than in the Danish elections.

Results comparing the Danish, Dutch, and American elections are shown in Figure 5.5. Note the clear clustering of each election system. The twelve American elections, shown in red, cluster in the bottom left hand corner near the origin, the eight Danish elections cluster in the center of the figure, and the four Dutch elections cluster on the far right. Using k-means clustering with 3 clusters only one example is misclassified (the outlying American example is classified as Dutch) resulting in a Rand index of 0.9289. From the scatter it is clear that the Danish and Dutch elections both usually have approximately the same cyclic component, but the Dutch elections consistently have a larger transitive component, and both have larger cyclic components than the American elections. The Danish and Dutch elections considered are both parliamentary elections among approximately ten parties, whereas the American elections are presidential elections among three to five candidates,

typically dominated by the pair of major party nominees. It is not surprising that in elections with more choices that voter opinion may be more heterogeneous, and as a result more cyclic (less well described by assigning a single rating to each party). These conclusions match the observation in [204] that political opinion in American presidential elections was more homogeneous than in other European examples (French and German).

The three systems are clearly separated in the size of their transitive component. The American presidential elections typically see a transitive component with size about 0.2, after normalizing by the number of edges in the network. The Danish elections typically have a transitive component of size 0.55, and the Dutch elections typically have a transitive component of size 0.9. Moreover, most of the spread within the American elections, and Danish elections, is along the transitive axis. The American election closest to the origin is 2016, and the outlying American election with transitive component near 0.6 is the 1968 election. The large transitive component in 1968 is a result of George Wallace's third party campaign against Nixon and Humphrey. Both Nixon and Humphrey have comparable ratings ($0.30 \pm 0.016, 0.18 \pm 0.015$ respectively) while Wallace has a significantly lower rating (-0.47 ± 0.017).[4]

The clear separation between the Danish, Dutch, and American elections in the size of their transitive component is reflected in the overall size of the estimated edge flows. In general, the larger in magnitude the log-odds edge flow, the farther the preference margins are from fifty percent. Therefore, the fact that the transitive component is larger in Dutch elections than Danish elections, and Danish elections than American elections, means that the average preference margins were larger in the Dutch elections than the Danish elections

[4]George Wallace was Governor of Alabama and ran on a pro-segregationist ticket. Wallace won multiple states in the Deep South, with the objective of preventing either Humphrey or Nixon from winning a majority in the electoral college, forcing a brokered convention [207]. While popular in the Deep South, and among blue-collar union workers, Wallace only won 13.5 percent of the popular vote, and only 8 percent in the North.

than the American elections.

The difference in preference margins is reflected in the best fit prior parameter to each system. For the American elections the best fit prior parameter is $\gamma = 5.67$, for the Danish elections the best fit prior parameter is, on average, 1.85, and for the Dutch elections is 0.97. Increasing the prior parameter, γ, decreases the probability of observing large win probabilities (large preference margins). If $\gamma = 1$ then all win probabilities (preference margins) are equally likely, while if γ is large win probabilities near one-half are more likely than win probabilities near zero or one. This difference in γ means that American presidential elections split voter preferences close to evenly, while Danish and Dutch elections do not. Thus the Danish and Dutch systems are better at maintaining minority parties which are only preferred by a small portion of the electorate. In and of itself this is not a surprise as both the Danish and Dutch elections considered are parliamentary, use proportional representation, and involve far more major parties than the American elections. These results underscore that the two-party nature of American politics prevents minority parties or candidates from competing successfully, and that the two major parties maintain support from close to half the electorate. Therefore the distinct separation of the three systems in their transitive components is a clear demonstration of how structural features of the different election systems is captured by the HHD.

A common feature of all the elections considered is that the transitive component is notably larger than the cyclic component. Therefore, while aggregate voter opinion cannot be adequately explained by an Elo type rating in all but a few of the American elections, most of the structure of the voter preferences can be described by rating the competitors. As discussed above, most of the variation in preferences between systems is also associated with the transitive component. This result supports our expectation that the elections would be mostly transitive, with a smaller, but nonzero, cyclic component.

In general we expect that the size of the cyclic component will be small relative to transitive component if the relative intransitivity (ratio of cyclic component to the overall edge flow) is small. The expected size of the the relative intransitivity is related to the correlation, ρ, in the log-odds that a randomly chosen voter prefers A to B, with the log-odds that another randomly chosen voter prefers A to C. The larger this correlation the more homogeneous voter opinion regarding candidate ranks. This correlation is $1/2$ at its largest, and 0 at its smallest, provided the trait-performance assumptions (see Section 4.6.2) hold. If the correlation is $1/2$ then competition is necessarily perfectly transitive. Therefore the correlation ρ is a statistical feature of the distribution of voter opinion, i.e. the culture, that is directly related to the expected sizes of the cyclic and transitive components.

We estimated ρ for each election using both point estimation, and sampling. The estimated correlation coefficient remains consistent across all eight Danish elections, ranging from 0.440 in 2001 to 0.471 in 1998. In addition, the predicted relative intransitivity using the correlation coefficient matched the estimated relative intransitivity for all eight elections. The correlation coefficient also remained consistent across the four Dutch elections, ranging from 0.477 in 1994 to 0.487 in 1989. Just as in the Danish examples the predicted size of the relative intransitivity using the correlation coefficient matched the actual size in all four elections. The fact that the Dutch elections have a larger correlation coefficient (coefficient close to $1/2$) corresponds to the observation that the Dutch elections have a smaller cyclic component relative to their transitive component than the Danish elections, so are relatively less cyclic. It follows that voter preferences was more homogeneous in the Dutch elections than the Danish elections. The correlation coefficient ρ varies much more in the American examples than the Dutch and Danish examples, ranging from 0.313 in 1980 to 0.499 in 1968.

Note that none of the observations made here would have been possible using a discrete

transitivity measure, since all of the elections considered were transitive. Therefore, all of the elections considered would be identical using a discrete measure.

The HHD: The Cyclic Component

Since most of the observed elections have a latent cyclic component it is natural to ask, which edges have the largest cyclic component? The edges with the largest cyclic component should correspond to the edges where the observed win frequencies are worst predicted by the competitor ratings. If we rank the edges in order of largest cyclic component (in magnitude), and compare that ranked order to the edges ranked by worst match to data using the perfectly transitive hypothesis, we find that the edges with a large cyclic component are the edges which are worst explained by the perfectly transitive hypothesis.

In the Danish elections the edges with the largest cyclic components consistently involved at least one of the minority parties who outperformed their rank using MR or the Hodge rating. For example, in all four elections between 1973 and 1979 the edge with the largest estimated cyclic component connected the Socialist People's Party and the Communist party. The same pattern appears in 1998 and 2001, when the largest estimated cyclic component appears on the edge between the Socialist People's Party and the Socialist Unity List party - the successor to the Communist Party. The same edge was the third most cyclic in 1994. Other minority parties (Danish People's Party, Progressive Party, Minority Party) appear frequently in the ranked list of most cyclic edges. In addition, the size and sign of the cyclic component on these edges remained consistent across the course of the eight elections. The consistent recurrence of the same edges in the list of highly cyclic edges, with approximately the same size component, suggests that these observed trends are a systematic component of Danish politics. These latent intransitivities may explain some of the differences observed in the election outcomes and the MR rankings.

For the American examples, we focus on the pilot studies performed in 2016 and 2019. In both cases, if only candidates from a single party are considered (Republican in 2016, Democrat in 2019), then the estimated cyclic component is small, and we accept the hypothesis that voter opinion is perfectly transitive. However, if candidates from both parties are considered (add Sanders and Clinton to the Republican candidates in 2016, or Trump to the Democrats in 2019), then the estimated cyclic component is about twice as large, and we reject the hypothesis that voter opinion is perfectly transitive. If Clinton and Sanders are added to the Reuplican candidates considered in the 2016 pilot then the four edges with the largest cyclic component all include at least Sanders and Clinton, and three of the four involve one Democrat and one Republican. The edges are, in order, Sanders over Trump, Sanders over Clinton, Clinton over Rubio, and Clinton over Carson. Similarly, if Trump is added to the list of Democratic candidates considered in the 2019 pilot then the two edges with the largest cyclic component both involve Trump (Harris over Trump, and Biden over Trump). Thus, voter opinions regarding candidates within a party are more transitive than voter opinions regarding candidates from both parties.

5.6 Social Hierarchies

5.6.1 Motivation

Many social animals engage in competition for dominance within social hierarchies. Examples include species from a wide range of taxa including primates, ungulates, birds, fish, and social insects [210, 119, 211, 212, 213, 214, 215, 216, 217, 95, 118]. Success in competition is associated with priority access to resources [218, 213, 219, 220], territoriality [221], and higher reproductive output [214] . For example, in a study of white-face

capuchins, Muniz [214] found that alpha-males produced 62 to 72 percent of all offspring within each group, and that most baby monkeys not sired by the alpha-male were born to his daughters, who do not mate with their father. This degree of monopolization is representative of a variety of primate species. As a consequence, success in competition events within social hierarchies can play an important role in evolution, and traits associated with success may be strongly selected for. The importance of success in these forms of competition has lead to nearly a century of prolonged study of animal hierarchies, both through empirical studies of wild and captive populations, and through behavioral ecology theory [222, 219]

Interest in social hierarchies among animals is usually traced back to Schjelderup-Ebbe [223] who coined the term “pecking-order" to describe the remarkably hierarchical behavior of domestic fowl [218]. In a pair of landmark papers Landau [117, 224] introduced a measure of linearity (transitivity), usually denoted h, to quantify the hierarchies documented by Schjelderup-Ebbe, and discussed how these hierarchies may arise from trait distributions and social factors. Landau famously concluded that the degree of hierarchy observed in Schjelderup-Ebbe’s hens could not be plausibly explained by null-models of trait distributions, and that other social factors must play a role in establishing and maintaining hierarchies. This narrative is the blue-print for nearly a century of similar studies in different species. Typically a group of individuals (either wild or captive) is monitored, outcomes of competitive or agonistic interactions are recorded, a measure of linearity (transitivity) is computed based on the recorded interactions, then the individuals are ranked and ranks are correlated against possible traits that may confer success. The degree of linearity is often correlated with features of the group studied (group size, density of competitive network), and the impact of other social factors are considered [119, 221, 220, 118, 216, 225].

Since Landau, quantitative measures of societal structure, in particular measures of linearity/transitivity, have played an essential role in the study of animal societies. The most widely used measure, h', is a variant of Landua's original measure proposed by de Vries [143], which is, in turn, a variant of Kendall's measure which is widely used in paired comparison [97]. Kendall's measure is based on a normalized count of the number of cyclic triangles in a complete tournament. Kendall's measure can be computed analytically by computing the variance in the number of competitors each competitor usually beats. Landau's measure is a normalized version of this variance, and is equivalent to Kendall's measure when the number of competitors is odd, and slightly smaller when the number of competitors is even [143, 95]. The difference in the two measures is associated with their normalization constants. Landau's measure is preferable when tied relationships are possible, and de Vries modified measure, h', is a variant of Landau's measure that allows for tied relationships, and incomplete tournaments where some pair of competitors either cannot, or are never observed, to compete [143]. The variant is computed by averaging the Landau measure over imputed data, where the dominance relation between pairs of competitors who could compete but are not observed to have competed is chosen randomly, uniformly, and independently. In [143] de Vries also proposed a randomization test, which is almost universally used to evaluate the statistical significance of an observed degree of linearity. All of these measures equal one if the observed system is transitive, and decrease the more cycles are present in the system.

Empirical studies of social hierarchies among animals are limited by the difficulty of gathering the necessary data [211]. Populations must be tracked, sometimes over large areas, must be habituated to observation, many individuals must be identified and monitored, and monitoring often must be sustained over long periods of time, in some cases spanning multiple decades (cf. [214, 213]). The difficulty and cost of gathering enough

data depends on the species and the scope of the study. Experimental populations are easier to study, but experimental populations may not behave the same as wild populations since access to resources is assumed to be an important factor influencing realized social structures [213, 219, 225]. Moreover, experiments and are often limited to small animals that can be easily kept in captivity (cf. [118]). In an experimental setting researchers may be able to control which pairs of individuals compete, and how often they compete, however care has to be taken to ensure that the animals are not injured by the competitive encounters which may limit the number of repetitions that can be ethically performed [174, 118]. In an observed population the schedule of competition events is set by the animals themselves. As a result, the number of repeated interactions per pair usually varies widely across pairs, with some pairs competing frequently, while other rarely if ever compete. This imbalance in the data can be a challenge when using methods derived for paired comparisons, where it is usually assumed that the number of repeated events per pair can be controlled. Individuals may avoid competing with other individuals if they expect to lose, or are at risk of injury during competition. Thus the number of repetitions per pair is not independent of individual ranks within the hierarchies, with more frequent interactions between individuals with similar rank [226, 215], and more interactions for individuals with intermediate ranks [213].

Often there are some pairs, if not many pairs, of individuals who never compete during the observation period as “even intense observational effort can never guarantee that most dyads within the group will be observed in agonisitic encounters" [212]. This problem is compounded by issues in choosing the number of individuals and length of time considered. The more individuals considered the more pairs do not compete [211], and observation periods should not be long enough that the structure of competition within the society changes significantly. When there are pairs who do not compete the investigator must

determine whether the pair could not compete, or could have competed but didn't during the observation period, since in the first case the "zero" is structural, and in the second case it means the investigator is missing data. Multiple studies [211, 212, 95] have shown that the standard measures of linearity, including h', decrease the more pairs are missing, and that the percent of pairs without observed encounters explains more of the variance in measured linearity than any other feature of the groups considered. Shizuka demonstrated that this trend is a result of the imputation procedure used to fill in missing data [95]. A more general version of the missing data problem is that there are often pairs of competitors who rarely compete. Often there are many pairs who only interacted a handful of times. This can make estimation difficult as sample sizes are often small on many edges.

Last, but not least, the ways in which individuals compete vary. Pairs often engage in a variety of agonistic interactions of varying severity, ranging from displays of aggression and submissiveness to physical conflict [218, 211, 213, 118, 225]. Therefore, even if the same pair of individuals competes multiple times they may not always compete in the same way, nor are the outcomes of repeated interactions necessarily independent. There is strong evidence to suggest that "winner" effects play a large role in animal societies, where the winner of a previous contest is favored in future contests. The outcome of competition between one pair may not even be independent of the outcome among another pair if the animals observe the outcome of the other event, or if animals form alliances and coalitions (cf. [214]). Thus the study of competition for dominance in social hierarchies among animals is uniquely challenging, and often requires highly system specific knowledge.

Despite these difficulties, study after study have show high degrees of linearity in different animal societies. This is often true despite strikingly low event counts. That said, while the index of linearity is usually large, it is also frequently not equal to one (cf. [213, 216]), indicating that at least some intransitive cycles are commonly observed

in animal societies [174, 221]. For example, Yasukawa observed a cycle among dark-eyed juncos in which individual B3 beat B4 in 22 out of 24 events, B4 beat B5 in 19 out of 23 events, but B5 beat B3 in 11 of 15 events [225]. Cycles like this may have arisen due to sampling error, however statistical significance with respect to the possibility that observed interactions do not reflect true dominance relations is largely missing from the literature. Alternatively, cycles may arise for structural reasons. An example is documented by Shoemaker who noted that male canaries typically dominant females, unless that female is their mate, and that these cycles may be intensified by breeding [221]. Combined these observations suggest that animal hierarchies are dominated by a transitive component, but often contain some cycles which may be either structural or incidental.

The standard approach when computing linearity for social hierarchies is to start by forming a dominance matrix [133, 143, 100]. The A, B entry of the dominance matrix is one if A beat B more often than B beat A, one-half if A beat B as many times as B beat A, and zero otherwise. Thus the standard approach discards any information about the frequency with which one competitor beats another except, and only considers whether a competitor wins a majority of the observed events. Competitor A is said to dominate B if A wins the majority of events observed against B. This methodology is motivated by another common trend observed in animal societies: observed win records are often extremely one-sided, with a single competitor in each pair winning almost all, if not all, events observed between the pair [174, 223]. This asymmetry is often true even when many events are observed between the pair. For example, a long-term study of social hierarchies among female mountain gorillas found that only 4 percent of all observed events were upsets [216]. A quick survey of 176 published win records made available by [95] confirms that most animal hierarchies are both strikingly linear, and strikingly asymmetrical.

The striking asymmetry of win records in animal competition events motivates the

concept of dominance, a “pattern of repeated, agonistic interactions between two individuals, characterized by a consistent outcome in favour of the same dyad member and a default yielding response of its opponent rather than escalation" [143, 218]. This sort of yielding behavior is widely documented in social species (cf. [214]) and may even occur spontaneously without an aggressive trigger. There are many reasons why a competitor might yield to another, not the least if escalation is dangerous or costly and they are unlikely to win [227, 225]. In this way the outcome of early events may determine the outcome of many subsequent events. Alternatively conflicts may be resolved by irrelevant but recognizable differences to avoid risky conflicts [227, 225].

A society is said to be despotic if dominance relations strongly predict competition outcomes, and a dominant individual rarely loses a competition event. A society is egalitarian if the outcome of competition is not well predicted by dominance relations [212]. Societies are characterized as despotic or egalitarian based on measures of steepness. A high steepness indicates win probabilities near zero or one, and a low steepness indicates win probabilities near one half [228]. If a society is despotic then each observed competition outcome carries more information than if a society is egalitarian. This motivates the standard approach in which an individual is considered dominant over another if they win a majority of events between the two, and allows the dominant individual in a pair to be identified even if only one event is observed. The fact that upsets do occur, even in highly despotic societies, suggests that the outcome of competition should still be considered a sample from a probabilistic event, however the usefulness of this perspective depends on how despotic the society is. Obviously, the closer the win probabilities are to one half the more information is discarded by rounding probabilities to zero, one half, or one. Moreover, when an individual has only won one more event than an opponent it is clear that denoting the individual as dominant may be prone to sampling errors. An important advantage

of considering despotic societies with steep dominance hierarchies is that by reducing win probabilities into a categorization of dominant and subordinant a dominance based approach can avoid introducing statistical assumptions about the competition events that may not be valid for animal competition [180, 228]. For example, the standard approach used in pairwise comparison, and in our statistical framework, is to assume that the outcome of distinct competitive events are independent, depend only on the pair of competitors competing, and the win probabilities are fixed over the duration over the study period. Any of these assumptions may be, and are, violated in some species [180, 228]. In this work we limit our attention to egalitarian societies since they are a better fit to our statistical assumptions. We will further limit our attention to systems with many interactions observed per pair so that we have large enough sample sizes to draw conclusions with some statistical significance. Of the 176 data sets considered in [95] these two criteria limit our attention to ten studies on six different captive species.

While the data available, and underlying statistical assumptions regarding social hierarchies in animal populations and politics are markedly different, some of the questions asked are noticeably similar. Both fields have shown a sustained interest in measuring the degree of linearity/transitivity present in a system, and in both fields there is strong empirical support for the hypothesis that competition is primarily transitive. In both fields not all systems studied are transitive, though the interpretation and implications of intransitivities differ. Moreover, comparison between studies is difficult in both fields because methodology differs [213, 170]. This problem is harder to resolve in animal societies since competition in animal societies, and the structure of animal societies, varies widely between, and sometimes within [213], species.

In both politics and animal societies issues of statistical significance of results are important, and though the significance of measures of societal structure are commonly

discussed in animal behavior, statistical significance is usually measured with respect to a null model [133, 143, 180], and does not account for possible sampling error.[5] In both fields the observation that systems are predominantly, but not entirely, transitive has prompted theory to explain why transitive structure appears so universal. Behavioral ecology theory attempts to explain the patterns of societal structure observed in animal populations from an evolutionary perspective [222, 219]. Other theory attempts to explain the prevalence of transitivity by considering the role of traits in mediating competition outcomes [117], and

[5]The standard randomization test used to test the significance of an observed degree of linearity was proposed by de Vries [143] as an improvement to the test proposed by Appleby [133]. Both tests are based on a significance test proposed by Kendall [97]. In the randomization test a series of random dominance matrices are sampled, the linearity index is computed for each sampled dominance matrix, and the fraction of sampled matrices that are more linear than the data is recorded. For a large enough sample size this fraction approximates the probability of sampling a dominance matrix that is at least as linear as the data from a null distribution of dominance matrices. The null distribution used is a uniform distribution over all dominance matrices. Thus the randomization test tests the significance of an observed degree of linearity relative to a uniform distribution of dominance matrices. In this null model all dominance matrices are equally likely, so the observed degree linearity is declared significant if it is more linear than a chosen fraction of dominance matrices of the same size. Therefore a degree of linearity is determined significant when it is possible to reject the null hypothesis that all dominance matrices are equally likely. Note that this is not the same as confirming the hypothesis that the society is transitive, nor does it confirm a hypothesis that the society is mostly transitive, with a limited degree of intransitivity. Instead it shows that the society has more transitive structure than is plausible if all dominance relations were random and independent. This would be equivalent to rejecting the impartial culture hypothesis in a political setting. Then the significance reported is not a confirmation of transitivity, only a confirmation that the society is more linear than a purely random society. This significance is only useful so far as we expect the uniform null model to be a plausible model for competition, or as it restricts the space of possible competition structures. The fact that study after study identifies significant linearity suggests that the uniform distribution is not a plausible model for competition in most cases, just as the impartial culture model is not plausible in most political settings. Rejecting the hypothesis that all dominance matrices are equally likely also fails to significantly limit the space of possible dominance distributions since it is entirely possible to imagine other distributions of dominance matrices with a higher average level of linearity without requiring that all of the dominance matrices are transitive. Thus the standard significance measure is evaluated with respect to a distribution that is something of a straw-man. If the desired hypothesis we wish to test is that the society is transitive, then without a statistical model for competition events to explain any observed intransitivities there is no way to test for the significance of the observed degree of linearity. Unless there is a statistical model that can be used to account for sampling error the observed dominance matrix is treated as truth, so significance must always be computed with respect to a null distribution of dominance matrices, not with respect to the possibility that there are errors in the observed dominance matrix. As a consequence very few studies of animal societies attempt to distinguish between incidental intransitivity and structural intransitivity. This is a significant limitation of the dominance based approach in which win probabilities are replaced with dominance relationships. That said, these limitations are inevitable if the complexity of the competition event prevents reasonable statistical modelling. The reasons why statistical modelling of competition events can be difficult are discussed in [180].

the possible influence of “winner" effects [224, 216].

Therefore we perform an analysis of competition between animals that parallels the analysis we performed for the political examples. First we test the hypothesis that each society is perfectly transitive, that is, that the win probabilities satisfy an Elo/Bradley-Terry type predictive rating. This is motivated by the observation that most animal societies are predominantly transitive, and follows similar efforts by [229]. Note that, unlike the tests for significance of linearity, this hypothesis test accounts for possible sampling error. Next we estimate the size of the cyclic component and transitive components using both point and interval estimation. This analysis provides a pair of continuous measures of the absolute strength of competition of both types, thus provides a complementary alternative to the discrete indices of linearity which measure the relative strengths of the two components. We expect that animal societies have a small cyclic component relative to their transitive component. Since the HHD also produces ratings of the competitors we compute a “steepness" measure. In particular we plot the estimated ratings in decreasing order and find the slope of the closest linear fit. This follows the steepness measure proposed by [230]. In addition we report the estimated parameter of the prior as measures of how egalitarian or despotic competition is. To conclude we estimate the correlation coefficient ρ. If success in competition is mediated by the traits of the competitors, and the traits are drawn i.i.d. from a trait distribution, then the size of ρ controls the expected relative sizes of the cyclic and transitive components (see Section 4.6).

Since most studies of animal hierarchies adopt a dominance approach it is important to highlight that, following [231, 229, 232, 233], we assume that the outcome of competition events are independent, depend only on the pair of competitors competing, and that the win probabilities are fixed for the duration of the study. Other maximum likelihood ranking methods used in social hierarchies have adopted these assumptions in order to test a range

of transitivity hypotheses (see [180] for a review). In addition we assume that no win probabilities equal one or zero.

5.6.2 Data

The examples analyzed in this paper were chosen from a collection of 176 data sets compiled by Shizuka in a meta-study of indices of linearity and imputation of missing data [95]. We consider 18 different data sets, from seven different studies, each on a different species. The systems were selected since they are sufficiently egalitarian, and have large sample sizes. They are all examples of studies on captive populations. When possible, examples were also chosen that included repeated trials.

The examples considered are, in chronological order of study date, Masure: white king pigeons and shell parakeets (2 trials each) [174, 234], Bennett: ring doves (3 trials) [235], Shoemaker: canaries [221], Yasukawa: dark eyed juncos (6 trials) [225], Nelissen: cichlids [215], Solberg: house sparrows (3 trials) [220]. Note that all but one of these examples involve competition between small birds.

5.6.3 Results

Despotism vs. Egalitarianism: Fitting for γ

In order to apply the Bayesian estimation scheme we need to fit the prior parameter γ. The size of the prior parameter controls how egalitarian or despotic we expect a system to be since it controls the prior distribution of win probabilities. The larger γ the more egalitarian a society. If $\gamma = 1$ then all win probabilities are equally likely (uniform prior), if $\gamma < 1$ then the prior is u-shaped, and if $\gamma > 1$ then the prior is maximized at $1/2$. We fit for γ per species, so that the value of γ is based on all available data on the given

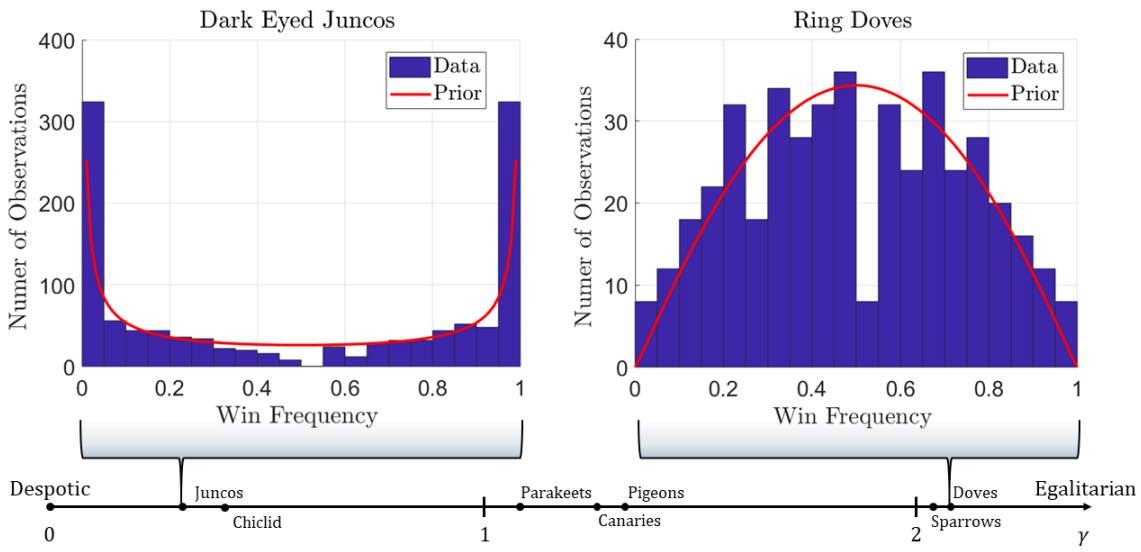

Figure 5.6: Priors for the most despotic species (Juncos) and egalitarian species (Doves) are shown, along with the distribution of observed win frequencies (histograms). The axis on the bottom represents the range of possible prior parameters. The values of the prior parameter for each species are marked. When $\gamma = 1$ the prior is uniform.

species. For example, we consider six different data sets published in [225] on dark eyed juncos. The prior parameter is fit to all of the six data sets, then used for each data set individually. Note that most of the studies considered single-sex populations separately, and the structure of competition among males and females differs in some species. We found that the estimated values of the prior parameter did not change significantly when fit to individual sexes relative to the variation between species, so used one prior for both sexes of a given species. This could easily be revised in future work. Estimates of γ in decreasing order of egalitarianism (increasing order of despotism) are as follows: ring doves 2.08, house sparrows 2.03, pigeons 1.33, canaries 1.27. parakeets 1.08, cichlids 0.40, juncos 0.30. The prior distributions for the most despotic, and most egalitarian species are shown in Figure 5.6. Note the marked difference in the shape of the prior distribution for the most despotic and most egalitarian species considered. Other species fall on the

continuum between these two examples. Parakeets, for example, have a nearly uniform prior distribution.

Hypothesis Testing

As in Section 5.5.3, we tested the hypotheses that the societies were perfectly transitive (win probabilities match an Elo type predictive rating), and perfectly cyclic. For most of the examples considered both hypotheses are rejected with sample p-values less than 10^{-3}. The exceptions are described here. An example from Bennett's study of ring doves [235] was close to plausibly perfectly cyclic with sample p-value of 0.049. This is a marginal rejection at best, however this was the first of three examples from Bennett's study considered. The first example was a study of male ring doves taken over the course of 24 days in the Summer. The remaining two examples were studies of female and male ring doves in the winter with observation lasting 54 days. Consequently, event counts were larger in both winter examples. In both cases the perfectly transitive hypothesis is rejected with sample p-values less than 10^{-3}. The estimated values of other parameters did not change much between winter and summer for the male ring doves, therefore we reject the hypothesis that ring dove societies are perfectly transitive with much higher confidence than is implied by the outcome of the first study. This is representative of other marginal cases (one of the two pigeon cases, two of the six junco cases). For example, one of the six dark eyed junco data sets passed the perfectly transitive hypothesis test with sample p-value 0.168, and another came close to passing the hypothesis test with sample p-value 0.041, while the remaining four tests are rejected with sample p-values less than 10^{-3}.

Solberg's study of house sparrows [220] is clear outlier in the hypothesis testing. The first data set comes close to passing the perfectly transitive hypothesis test (sample p-value 0.003), and passes the perfectly cyclic hypothesis (sample p-value 0.081). The

second and third both pass the perfectly transitive hypothesis test (p-values 0.639 and 0.573 respectively), and the second also passes the perfectly cyclic test (p-value 0.914). It is impossible for a system to be perfectly cyclic and perfectly transitive unless all the win probabilities equal one-half. Therefore this indicates that the house sparrow examples do not contain enough observed events to draw confident conclusions regarding the type of competition observed (the second house sparrow example only averages 8.21 events per pair of sparrows). These examples are included to show why uncertainty quantification is important when estimating the sizes of the components, how we detect bias due to uncertainty, and ultimately reject data sets with insufficient sample sizes.

The HHD: Measures and Characterization

Next we estimated the sizes of the transitive and cyclic components. Point estimates using the conditional expected edge flows are reported in Table 5.4. Uncertainty in the point estimate is reported (one standard deviation in posterior), along with credible intervals, percent bias that could arise from sampling error, and the credible bounds. The credible interval is derived by finding highest posterior density interval (HPDI) containing 95 percent of the posterior probability. It does not necessarily contain the point estimate since the point estimate is derived by applying the measure directly to the estimated flow, rather than finding the posterior for the measure and then estimating the measure based on its posterior. This is done to reduce the bias due to sampling error and uncertainty (see 5.4.1 and the supplement for discussion). The expected percent of the estimated measure squared contributed by sampling error is reported for each measure. Note that the actual percent of the measure contributed by noise is unknowable, so these percents are provided primarily to give a gauge for how much the point estimator and credible interval might be influenced by sampling error. When this percent is large then both the estimate and the credible interval

		Transitive				Cyclic			
Example	# Events	Estimate	Credible Interval	% Bias	Credible Bounds	Estimate	Credible Interval	% Bias	Credible Bounds
Canaries	237.6	0.54 ± 0.02	$[0.51, 0.59]$	1.1	$[0.39, 0.84]$	0.48 ± 0.02	$[0.47, 0.53]$	5.5	$[0.36, 0.85]$
Doves 1	17.4	0.28 ± 0.04	$[0.25, 0.40]$	21.2	$[0.04, 0.84]$	0.26 ± 0.05	$[0.30, 0.48]$	77.8	$[0.001, *]$
Doves 2	157.3	0.35 ± 0.03	$[0.30, 0.41]$	2.8	$[0.22, 0.50]$	0.29 ± 0.03	$[0.25, 0.36]$	6.0	$[0.16, 0.45]$
Doves 3	82.7	0.60 ± 0.05	$[0.53, 0.70]$	2.3	$[0.43, 0.90]$	0.39 ± 0.04	$[0.35, 0.48]$	8.1	$[0.25, 0.61]$
Cichlid	39.1	1.28 ± 0.11	$[1.26, 1.62]$	2.7	$[0.84, 2.06]$	0.66 ± 0.11	$[0.63, 0.98]$	26.8	$[0.21, 1.78]$
Pigeons 1	70.2	0.59 ± 0.04	$[0.55, 0.69]$	2.5	$[0.41, 1.45]$	0.21 ± 0.05	$[0.19, 0.36]$	47.2	$[0.001, 1.58]$
Pigeons 2	60.1	0.64 ± 0.08	$[0.59, 0.83]$	3.4	$[0.39, *]$	0.30 ± 0.08	$[0.28, 0.53]$	37.4	$[0.08, *]$
Parakeets 1	21.2	0.73 ± 0.11	$[0.69, 1.06]$	7.0	$[0.36, *]$	0.51 ± 0.10	$[0.52, 0.83]$	39.0	$[0.22, *]$
Parakeets 2	30.2	0.80 ± 0.10	$[0.75, 1.09]$	5.1	$[0.48, *]$	0.35 ± 0.09	$[0.36, 0.65]$	65.8	$[0.07, 1.65]$
Juncos 1	36.5	1.46 ± 0.33	$[1.46, 2.51]$	7.3	$[0.85, *]$	0.40 ± 0.36	$[0.44, 1.59]$	132.8	$[0.000, *]$
Juncos 2	36.1	1.31 ± 0.33	$[1.33, 2.40]$	10.2	$[0.75, *]$	0.54 ± 0.36	$[0.62, 1.81]$	79.2	$[0.01, *]$
Juncos 3	54.2	1.03 ± 0.18	$[0.97, 1.55]$	7.1	$[0.68, *]$	0.50 ± 0.30	$[0.44, 1.38]$	38.8	$[0.09, *]$
Juncos 4	43.9	1.14 ± 0.29	$[1.10, 2.02]$	8.3	$[0.65, *]$	0.78 ± 0.30	$[0.78, 1.72]$	34.8	$[0.41, *]$
Juncos 5	56.4	1.53 ± 0.33	$[1.56, 2.63]$	8.4	$[0.97, *]$	0.80 ± 0.35	$[0.81, 1.93]$	58.1	$[0.30, *]$
Juncos 6	50.1	1.47 ± 0.37	$[1.50, 2.68]$	7.9	$[0.85, *]$	0.62 ± 0.36	$[0.68, 1.83]$	76.8	$[0.21, *]$
Sparrows 1	12.4	0.29 ± 0.06	$[0.25, 0.45]$	24.0	$[0.000, *]$	0.31 ± 0.06	$[0.36, 0.55]$	69.6	$[0.05, *]$
Sparrows 2	8.2	0.21 ± 0.07	$[0.19, 0.42]$	60.2	$[0.000, *]$	0.22 ± 0.06	$[0.31, 0.52]$	176.4	$[0.000, *]$
Sparrows 3	22.2	0.49 ± 0.12	$[0.36, 0.75]$	12.7	$[0.20, 1.49]$	0.13 ± 0.08	$[0.09, 0.36]$	185.1	$[0.000, 0.86]$

Table 5.4: Estimated sizes of the transitive and cyclic components of the edge flow (normalized by $\sqrt{E}$) for 18 animal societies. The standard deviation in the posterior for each component is reported next to each estimated value. The credible interval is the 95 percent credible interval approximated by sampling from the posterior. The percent bias is the expected percent of the estimated measures (squared) associated with variance in sampling a win record. It should be emphasized that this is not the actual percent bias in the estimate, since the actual bias is unknowable. When this percent is large the credible bounds should be used for interval estimation, since the credible interval depends of the posterior for each measure, whose mean experiences approximately twice the bias as the estimator. The credible bounds are the smallest upper bound, and largest lower bound, for which the hypothesis that the log-odds satisfy the constraint is accepted with significance 0.05. This interval is always wider than the credible interval as it only requires that the MLE model satisfying the bounds is credible, not that the region inside the bounds contains most of the mass of the posterior distribution. A $*$ indicates that no bound was found. Note that we can often estimate the size of the transitive component more robustly than the size of the cyclic component, and can find lower bounds on the measures more often than upper bounds. This latter asymmetry is a result of the skew in the posterior when there are pairs of competitors for which only one competitor wins all events.

may be significantly biased by sampling error and uncertainty in the posterior. These biases motivate the credible bounds, which are based on hypothesis testing rather than the posterior distribution. The credible bounds are, respectively, the smallest upper bound, and largest lower bound, such that the region of edge flows defined by the bound passes the hypothesis test. The region outside of the credible bounds would not pass the hypothesis test and the value of the bounds are the smallest, and largest, that pass the hypothesis test. The species in the table are ordered in approximately increasing order of uncertainty.

First, note that the uncertainty in the estimates is generally much larger for the animal examples than for the political examples (see Table 5.3.) This reflects the differences in sample sizes. The political examples each average over a thousand respondents per pair of candidates, while the animal examples typically average 20 to 100 interactions per pair. Certain examples buck this trend, for example, the second dove study averages 157.3 interactions per pair and the canary example averages 237.6 examples per pair. As a result we can estimate the measures with the most certainty for these examples. In contrast, the house sparrow examples average only 12.4, 8.2, and 22.1 interactions per pair, and, as a consequence, do not allow for confident estimation.

Next, note that the transitive component can be estimated with systematically less bias than the cyclic component even though the variance in the posterior for both components is about the same. This is a result of the sizes of the cyclic and transitive subspaces. Since the cyclic subspace is typically larger than the transitive subspace a larger fraction of the uncertainty in the edge flow becomes uncertainty in the cyclic component. Moreover, since the estimated transitive component is typically larger than the estimated cyclic component the percent bias in the cyclic component is usually greater than the percent bias in the transitive component (more uncertainty in a smaller quantity).

For some of the animal examples the possible biases are small and the point estimators

are reliable. The canary example and second pair of dove examples all have percent biases beneath 10 percent for both measures, and have reasonably tight credible intervals and bounds that contain the point estimate. The first dove example has considerably more uncertainty, but, as discussed before, this difference in uncertainty is a result of the shorter observation period [235] used for the first dove example. Six more examples have moderate biases in the cyclic component, and have credible intervals containing the point estimate (the cichlid example, both pigeon examples, the first parakeet example, and the third and fourth junco examples). Therefore, half of the examples have sufficiently large sample sizes to ensure that either the point estimator or credible interval for the cyclic component are, at least reasonably, reliable. When the expected bias due to uncertainty is large then the credible bounds should be used instead of the credible interval. Note that we can often find a credible lower bound on the size of the cyclic component, even if the possible bias in the Bayesian estimator is large (see the last three junco examples). Since the bounds are based on hypothesis testing, if the lower bound on the cyclic component is greater than zero then the system will not pass the perfectly transitive hypothesis test. In contrast, if the bound equals zero (see the first junco example and last two sparrow examples) then the system could be perfectly transitive. When the bound is close to zero (see the first dove example, first parakeet example, or second junco example) then the perfectly transitive hypothesis is rejected, but without confidence (sample p-value close to the desired significance). Thus, even if there is too much uncertainty in the posterior to use the Bayesian estimators reliably, the bounds on the sizes of the components derived based on hypothesis testing are still informative.

In some cases we cannot find an upper bound on the size of a component. This occurs if there are many pairs of competitors for which a single competitor wins all observed events thereby skewing the posterior (see Section 5.4.1). For all eleven examples for which

we cannot find upper bounds on the sizes of the components at least ten percent of pairs include a competitor who has never won against their opponent. Typically these are low ranked competitors who are dominated by most of the other individuals.

For comparison, the largest lower bound on the cyclic component observed in all of the political examples was 0.13, and the largest estimated cyclic component across the political examples was 0.18. Eight of the eighteen animal examples have lower bound greater than 0.13, and seven have lower bound greater than 0.18. Therefore, despite the considerable uncertainty in the animal examples, close to half of the animal societies are demonstrably more cyclic than all of the political examples studied. Moreover, all three animal examples with small biases in the point estimator, have estimated cyclic component larger than any of the political examples, larger than the upper limit of the credible interval in any of the political examples, and larger still than the credible upper bound on the cyclic component in all but one of the forty political examples.

The size of the transitive component is also clearly larger in some of the animal examples than the political examples, and clearly smaller in others. In general, for the despotic societies, the overall size of the estimated edge flow is large (see the cichlid and junco examples), since the estimated win probabilities are often close to zero or one, and the transitive component is correspondingly large. In an egalitarian society (see doves or sparrows), the estimated edge flow is smaller since the estimated win probabilities are closer to a half, and the transitive component is smaller than in the European examples. Despite the potentially large biases in the estimated cyclic components, there is more variance in the estimated transitive components than cyclic components across all eighteen animal examples (see Figure 5.8). This mirrors the observation that most of the variance between political systems, and within political systems, was observed in the size of the transitive component not the size of the cyclic component. Thus more of the difference in

Example	β	Steepness	Credible Interval	R^2
Juncos 1	0.30	−0.41	[−0.76, −0.07]	0.73
Juncos 2	0.30	−0.53	[−0.95, −0.11]	0.75
Juncos 3	0.30	−0.46	[−0.58, −0.34]	0.97
Juncos 4	0.30	−0.51	[−0.79, −0.21]	0.97
Juncos 5	0.30	−0.71	[−1.00, −0.41]	0.91
Juncos 6	0.30	−0.56	[−0.97, −0.15]	0.78
Cichlid	0.40	−0.45	[−0.54, −0.35]	0.97
Parakeets 1	1.08	−0.22	[−0.35, −0.08]	0.78
Parakeets 2	1.08	−0.26	[−0.34, −0.18]	0.93
Canaries	1.27	−0.13	[−0.14, −0.11]	0.98
Pigeons 1	1.33	−0.21	[−0.30, −0.12]	0.88
Pigeons 2	1.33	−0.18	[−0.30, −0.05]	0.72
Sparrows 1	2.03	−0.08	[−0.10, −0.06]	0.93
Sparrows 2	2.03	−0.06	[−0.07, −0.05]	0.96
Sparrows 3	2.03	−0.28	[−0.44, −0.12]	0.97
Doves 1	2.08	−0.07	[−0.10, −0.05]	0.87
Doves 2	2.08	−0.24	[−0.44, −0.04]	0.82
Doves 3	2.08	−0.15	[−0.24, −0.07]	0.90

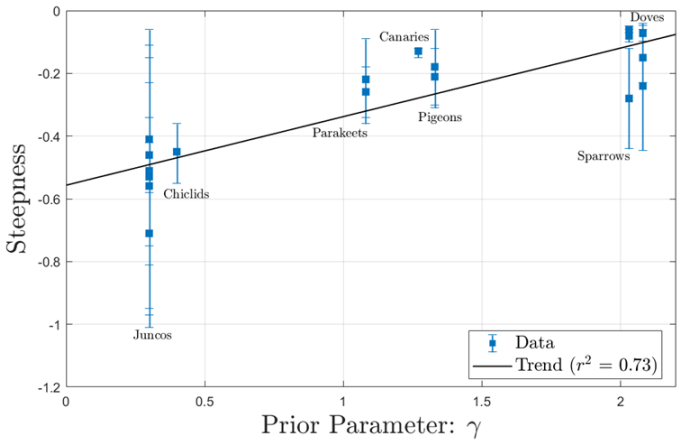

Figure 5.7: Steepness for the eighteen animal examples. The left panel tabulates the steepness for each example, along with the prior parameter γ, the credible interval for the steepness coefficient, and the R^2 value for the fit. The right panel plots the steepness against the prior parameter, and shows that there is a weak positive correlation between the prior parameter and the steepness.

the over-all size of the edge flow, and associated egalitarianism vs. despotism of a society, is explained by differences in the size of the transitive component, not the cyclic component.

Steepness and Despotism

This observation suggests that despotic animal societies are despotic due to large differences in the ratings of individuals. Steepness measures [230] measure how strongly win probabilities depend on differences in rank, and are generally computed by finding a best fit line to a set of ratings ranked in decreasing order. We fit for steepness by ordering the estimated Hodge ratings in decreasing order, then fitting for the slope of the best fit regression line through the estimated ratings (conditional expectation). Least squares regression is used with weights set to one over the variance in the posterior distribution for each rating. Steepness values, along with confidence intervals and r-squared values are reported in Figure 5.7, in order from most egalitarian to most despotic. The steepness controls how the odds that one competitor beats another depends on the difference in their

ranks. For every increase in rank the odds that the lower competitor beats the higher rank competitor change by a factor of $\exp(\text{steepness})$.

There is a weak positive correlation between the prior parameter γ and the steepness, indicating that the more despotic a society the steeper the ratings. That said, the ratings often do not fit well to a line. In some cases the high and low ratings form plateaus, with a plateau of high rated competitors and a plateau of low rated competitors. In others the competitors with intermediate rank all have similar ratings while the highest and lowest rated competitors have outlying ratings. This structure would be expected if ratings are distributed according to a bell-shaped distribution. Most commonly, the lowest rated competitor is much lower rated than the rest of the competitors. Not all of the linear fits are bad, for example the cichlid and canary examples both have ratings that fit well to a line. However, in general it does not appear that ratings are necessarily linear in the rank order, so rank difference is not a good predictor of performance of two competitors.

Cyclic Competition

Figure 5.8 compares the estimated sizes of the two components of the animal examples and the political examples. Note the overall upward shift in the scatter points for the animal examples. This reflects the larger estimated cyclic components, and while some of this shift may be attributed to greater uncertainty, as demonstrated before, much of the shift cannot be plausibly explained by bias due to uncertainty.

Unlike the political examples, which were all transitive, there are clear cycles in the animal data sets. In the first parakeet data set (female parakeets) there is a clear cycle between individuals labelled B, R, and 3. B beat R 16 out of 16 events, R beat 3 20 out of 26 events, and 3 beat B 21 out of 28 events. This triangle has the largest vorticity of any of the triangles among the competitors with vorticity 1.08, and the edge from 3 to B is the

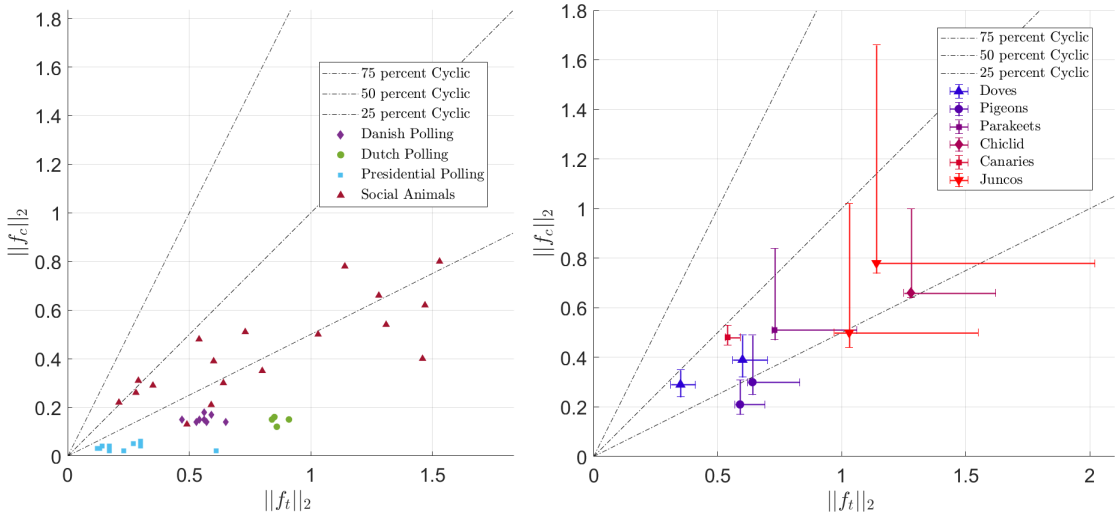

Figure 5.8: The left hand panel compares the estimated sizes of the components for the eighteen animal examples and twenty four political examples. The animal examples are marked with maroon triangles. The right hand panel shows the nine of the eighteen animal examples with moderate or small biases due to noise. The error bars denote the credible intervals, and marker type and color denotes the species. Note that the more egalitarian species (doves and pigeons) have smaller components, and the more despotic species (juncos and cichlids) have larger components. Also note the strong positive skew in the error bars. This reflects the skew in the posterior distribution.

most cyclic edge in the network. The next largest vorticity is 0.535 on the competitors B, O, and P. B beat P 22 out of 25 events, and P beat O 31 out of 34 events. Thus we would expect B to be far superior to O. Instead B only beats O 8 out of 15 events. Reversing the outcome of a single event between B and O would make this cycle intransitive. This example highlights that the cyclic component accounts not only for explicit cycles, but also for triangles where the win redocrds cannot be plausibly explained by a predictive rating. This triangle fails the “strong stochastic transitivity" hypothesis that, if $p_{ij} > 1/2$ and $p_{jk} > 1/2$ then $p_{ik} > \max\{p_{ij}, p_{jk}\}$ [231, 180]. The edges from B to P, and P to O are the second and third most cyclic edges in the graph, followed by the edges from R to 3, and 3 to B which appeared in the original cycle discussed.

Both of these triangles have vorticities that are clearly larger than the vorticities on the

remaining triangles. The next largest vorticities are 0.248, 0.208, and 0.131, all involving B. Of these top five cycles two involve the edge from B to O, three involve the edge from B to 3, and two involve the edge from B to Y. Thus B, who ranks in the middle of the flock when the flock is ranked by total number of individuals dominated [234], is involved in most of the strong cycles observed. The cycle between B, 3, and R, has a strong impact on the rankings when ranking by total number of individuals dominated. If we rank by the total number of individuals dominated then the rank order is 3, P, R, B, O, G, Y where 3, P, and R all dominated four other birds, B dominated three, O dominated three birds, and Y dominated none and won very few contests. Note that B, which is ranked beneath P and R by the total number of birds dominated, won the majority of contests with P and R by large margins (22 out of 25) and (16 out of 16). If the birds are ranked using the Hodge ratings then B is ranked first instead of fourth, P is ranked second, followed by 3, then O, then R who has fallen from the top three to the bottom three, then G, then Y. When accounting for the degree to which one bird dominated another, B is ranked highest, 3 is ranked third, and R is ranked fifth, but just counting birds dominated 3 is ranked first, R third and B fourth. Thus the strong cycle observed between B, 3, and R, has a strong influence on the rankings when ranking only by number of birds dominated and dominance relations. These two cycles are illustrated in Figure 5.9.

Other cycles worth highlighting occured in the juncos and the canaries examples. Two clear cycles are apparent in the junco examples, one involving three male juncos in which individual B3 beat B4 22 out of 24 contests, B4 beat B5 19 out of 23 contests, and B5 beat B3 14 out of 15 contests. The reversal of the relation between B5 and B3 is all the more striking because all other relations in the set of six birds are transitive, and 11 of all 15 pairs of birds had entirely despotic relationships, with a single individual winning all contests. If analyzed in isolation this cycle has vorticity 0.79, has relative intransitivity 0.95, and is

Example	# Reversals/Edge	Linearity: h	Pr{transitive}
Canaries	1.624/45	0.748 ± 0.067	0.0000^*
Doves 1	6.06/36	0.595 ± 0.127	0.0002^*
Doves 2	0.021/10	0.600 ± 0.009	0.0002^*
Doves 3	0.021/10	0.600 ± 0.009	0.101
Parakeets 1	0.753/21	0.642 ± 0.090	0.0000^*
Parakeets 2	1.593/21	0.874 ± 0.087	0.173
Sparrows 1	6.227/35	not complete	0.0000^*
Sparrows 2	8.220/33	not complete	0.0000^*
Sparrows 3	0.673/6	0.916 ± 0.178	0.805
Pigeons 1	1.579/21	0.853 ± 0.057	0.058
Pigeons 2	1.244/21	0.923 ± 0.070	0.270
Juncos 1	0.237/15	0.974 ± 0.055	0.797
Juncos 2	0.198/15	0.990 ± 0.034	0.920
Juncos 3	0.443/15	0.973 ± 0.082	0.875
Juncos 4	0.372/15	0.929 ± 0.157	0.821
Juncos 5	0.005/15	0.909 ± 0.012	0.997
Juncos 6	0.067/15	0.985 ± 0.057	0.936
Juncos 7	0.032/3	0.032 ± 0.176	0.032
Juncos 8	0.511/6	0.207 ± 0.066	0.006
Cichlid	0.0004^*/21	1.000 ± 0.003	0.9997^*

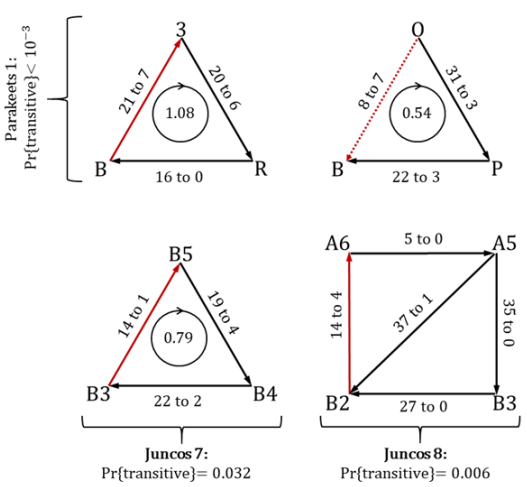

Figure 5.9: Cycles observed in animal populations and transitivity. The table in the left hand panel gives the average number of dominance relations that are flipped when sampling from the posterior per edge, the expected value of Landau's linearity index h over the posterior, and the posterior probability that the society is transitive. The rows are organized in order of increasing probability of transitivity. The four examples containing cycles discussed in the text are bolded. All quantities were estimated using 10^6 samples from the posterior, so quantities whose values are reported past the thousandths place are marked with an asterisk. Note that the first pair of dove examples, the first parakeet example, canary example, first two sparrow examples, and last junco examples are clearly intransitive, with posterior probability of transitivity less than 10^{-3}. The last dove example, parakeet example, and second to last junco example are possibly transitive, but with a small posterior probability. The cichlid example is very clearly transitive, and the remaining junco and sparrow examples are likely transitive. The right panel shows four of the cycles discussed in the text (two from the parakeets example, and the two additional junco examples). Arrows point from loser to winner. The red edge denotes the edge that points in the opposite direction predicted by heirarchy, and the dotted edge represents in the B, O, P cycle represents the edge that is far closer to neutral than expected. The vorticity associated with each triangular cycle shown inside each triangle, and the posterior probability that the corresponding systems are transitive is reported next to the brackets.

plausibly perfectly cyclic (sample p-value 0.51). A similarly noteworthy cycle is apparent in a different data set, this time involving four male juncos. Individual B2 beat B3 27 out of 27 events, B3 beat A5 35 out of 35 events, and individual A5 beat A6 5 out of 5 events. All three of these pairs have entirely despotic relationships. Nevertheless, A6 beat

B2 14 out of 18 events, forming a cycle between the four birds. Individual B2 also beat individual A5 37 out of 38 events forming a cycle between B2, A5, and A6. If analyzed in isolation this set of four birds has cyclic component of size 1.38, which is the largest cyclic component measured out of any of the data sets and has relative intransitivity 0.84. These examples clearly illustrate that there are strongly cyclic subsets of individuals within the junco examples, however, it is important to note that there are 90 triangles among the first six junco examples alone, so the significance levels for determining whether an observed cycle may be incidental should be chosen to reflect the number of triangles considered. The posterior probability that each example is transitive, and any observed cycles are incidental, is reported in Figure 5.9.

Shoemaker [221] discusses a series of nine cyclic triangles among ten canaries. Of the nine triangles, seven are produced by clear structural reasons. Male canaries dominate female canaries in almost all cases, except for their mate. In a mated pair the female dominates the male. This produces cycles when a male who dominates another male who dominates the first male's mate, is in turn dominated by his mate, or when a male dominates a female who dominates his mate. For example, individuals 15 and 55 were mates, and individuals 19 and 97 were mates [221]. The edges between 15 and 55, and 19 and 97, have a large cyclic component, and are the second and third most cyclic edges of all forty-five edges in the network. The triangles identified by Shoemaker account for the relatively high degree of intransitivity in the canary example, which is the most relatively intransitive of all examples considered.

Comparison to Political Examples

The animal examples are clearly more scattered than the political examples. This is true both within species (see right-hand panel of Figure 5.8), and between species (see left-hand

panel of Figure 5.8). The animal examples show much more variability within species than the political examples showed within nationality. This is not surprising as, there is more uncertainty in the animal examples and animal societies have noticeably large variations, both between species, and within species (cf. [213, 214]). Societies may evolve differently depending on access to resources [222, 219], and as a result animal societies are observed to vary in rates of agonism [236], intensity of conflict, degree of despotism, and linearity. Despite the larger variation within species, the results from the nine examples with moderate to low uncertainty show that societies drawn from the same species still tend to cluster (see the doves, pigeons, and juncos in Figure 5.8). This observation should be tested on more species, and with more than two trials for each species.

The significance of an observed index of linearity is usually evaluated by estimating the probability of sampling a random set of dominance relations that is equally or more linear. The dominance relations are usually sampled uniformly and independently [143, 230]. This sort of randomization test is widely used to evaluate the significance of measures of social structure (cf. [216]). We evaluate the significance of the estimated sizes of the transitive and cyclic components by sampling random win probabilities independently from the prior distribution, and then estimating the probability that, had the win probabilities been drawn from the prior, that we would have estimated a larger or equally large transitive component, or a smaller or equally small cyclic component. This tests whether the true win probabilities could plausibly have been sampled independently from the prior given the observed degree of hierarchy (large transitive component, small cyclic component). Ten of the eighteen animal examples had a significantly large transitive component, and thirteen had a significantly small cyclic component (significance 0.05). Two of the five examples with a non-significantly small cyclic component and two of the eight examples with non-significantly large transitive component were sparrow examples, which had low samples

sizes. All of the nine examples with large sample sizes were significantly less cyclic than would be expected had the win probabilities been sampled independently from the prior, except for two ring dove examples. Therefore, for most of the examples considered the examples were significantly more linear than would be expected if the win probabilities were sampled independently from the prior. Of the examples with large sample sizes only one species, the ring dove, could plausibly have win probabilities sampled independently.

The size of the cyclic component relative to the transitive component is larger for almost all of the animal examples than the political examples, and is larger for all nine animal examples with moderate uncertainty. This difference is shown clearly by the left-hand panel of 5.8. Even including the credible intervals, the animal examples are concentrated between the dashed lines demarking 25 percent and 50 percent cyclic. All of the political examples are concentrated beneath the 25 percent cyclic line. The relative intransitivity, $||f_c||_2/||f||_2$ of the ring doves, the least relatively intransitive animal studied, is between 0.53 and 0.67, or 0.56 and 0.59 depending on the example considered. In contrast the most relatively intransitive political examples have relative intransitivity between 0.30 and 0.32 (Danish), 0.18 and 0.21 (Dutch), and 0.18 and 0.59 (American). Of the nine animal examples shown in Figure 5.8 the canaries are the most relatively intransitive, with relative intransitivity between 0.65 and 0.67. Significance of the estimated relative intransitivities were computed by sampling win probabilities independently from the prior. All of the eighteen animal examples had significantly small relative intransitivities except for the two dove examples discussed before, and one of the sparrow examples. In many cases the observed degree of relative intransitivity was significant, with sample p-values on the order of 10^{-4} for the pigeons, cichlids, and canaries, and 10^{-3} for the parakeets and two-thirds of the juncos. Thus, for all species except the doves and sparrows we can reject the hypothesis that the win probabilities on each pair are independent, and for both the doves and sparrows

it is unlikely, but possible, that the win probabilities are independent. Therefore, for all the other species there is some correlation structure in the win probabilities that accounts for the relative transitivity of the social hierarchies.

As for the political examples we estimated the size of the correlation ρ. When the tournament is complete the estimated correlation coefficient ρ determines the estimated relative intransitivity, so all examples that were significantly less intransitive than would be expected if win probabilities were independent had significant correlation coefficients. The estimated correlation coefficients ranged from 0.155 (canaries) to 0.448 (juncos). In general the animal examples had more variation in correlation coefficients both within and between species than the political examples, and had smaller correlation coefficients on average. This matches the observation that the animal examples had a larger cyclic component relative to their transitive component.

In summary, the animal examples were, for most part, not perfectly transitive, had larger cyclic components than the political examples (both relative to the transitive component, and absolutely), and lower correlations. Like the political examples the animal examples showed more transitive structure than cyclic structure, with relative intransitivity significantly smaller than what would be expected if the win probabilities were all independent. Specific cycles, and edges with a large cyclic component, can be identified, and in some cases are explained by structural properties of the pair relations [221]. Also like the political examples the animal examples varied more in the size of the transitive component, and this variation accounted for most, but not all, of the difference between egalitarian and despotic societies. The animal examples showed considerably more variation in structure both within and between species than is observed in the political examples. Finally, smaller sample sizes in the animal examples require more careful analysis. By using bias estimates, we could identify examples with dangerously low sample sizes, and by using hypothesis

testing to compute credible bounds we could draw some, albeit weaker, conclusions about systems with low sample sizes.

5.7 Sports

Sports are a natural application area for the techniques developed in this chapter. Sports also present a challenging estimation problem since sports teams rarely play many repeated games per pair. Major League Baseball (MLB) is a promising exception, as MLB teams play many games per season, and 19 games per pair within a division.

We collected historical win/loss data for every year of MLB history since 1880, and analyzed data from each season, with prior models fit to decade and league. For comparison win/loss records from football and basketball were collected and analyzed. All data was gathered from FiveThirtyEight. The prior distribution distribution was estimated for each ten year interval since 1880, and the estimation procedure for the HHD was applied to the 1999-2019 seasons.

Despite the moderately large number of games played per pair of teams within division, the results were almost entirely inconclusive because the best fit for the prior parameter, γ, ranged from 21 to 26 over the course of the twenty years considered (fit to a ten year sliding window). Results for baseball, basketball, and football are shown in Figure 5.10. Since γ is large for baseball, the prior distribution of win probabilities for baseball is tightly distributed about one-half. As a result, we expect most baseball teams to be close to evenly matched, and require large win margins before estimating large win probabilities. It would require about 20 games per pair before the resulting estimates would be more informed by the data on the given pair, than by our prior expectation that baseball teams are evenly matched. Since the prior introduces a conservative bias, the resulting estimates

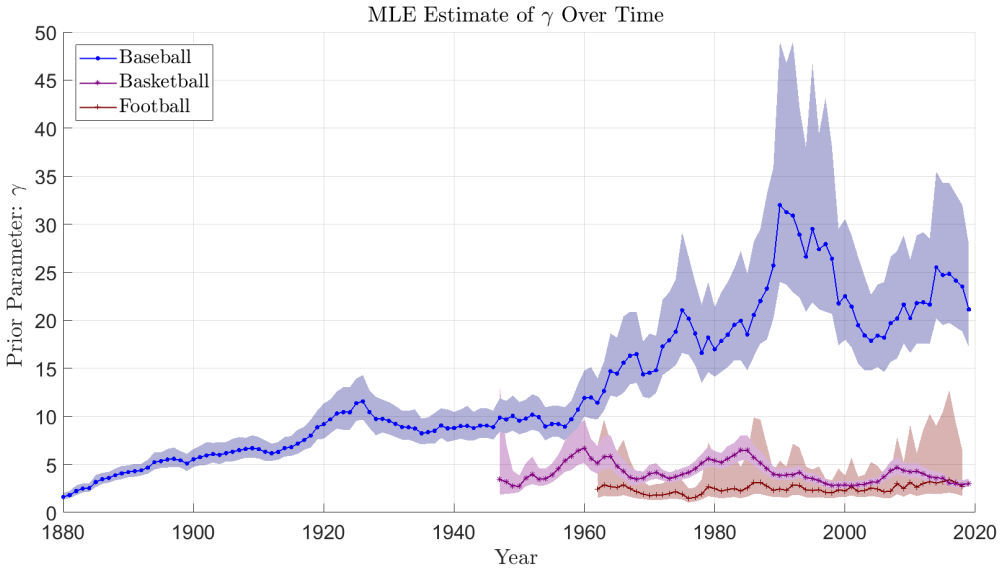

Figure 5.10: MLE estimate of the prior parameter γ on a ten year sliding interval for professional baseball, basketball, and football. Marked lines represent MLE estimates for γ, and the shaded interval represents the 95 percent HPDI. Results for baseball are shown in blue, basketball in purple, and football in red. Note the marked difference in the size of the prior parameter in baseball than in basketball or football. The larger γ the more concentrated the prior distribution about fifty percent win probabilities. The value of the prior is an effective number of games, and unless more games are observed than γ the estimation depends more on the prior distribution than the data available for each pair of teams.

for the log-odds are all small. Thus, even though the large (relative to other sports) event counts in baseball reduce the variance in the posterior, the signal is also small since most baseball teams are evenly matched. As a result the analysis returns inconclusive results, with wide confidence intervals, moderate to large p-values, and large estimated biases due to uncertainty. Thus, unlike in the political and animal examples discussed, we cannot discern whether cycles observed in baseball are structural or incidental.

The estimation framework developed in this chapter can be extended to incorporating more game data than game outcomes. The framework developed in this chapter models and estimates of the win probabilities using only game outcomes. Historical records of game

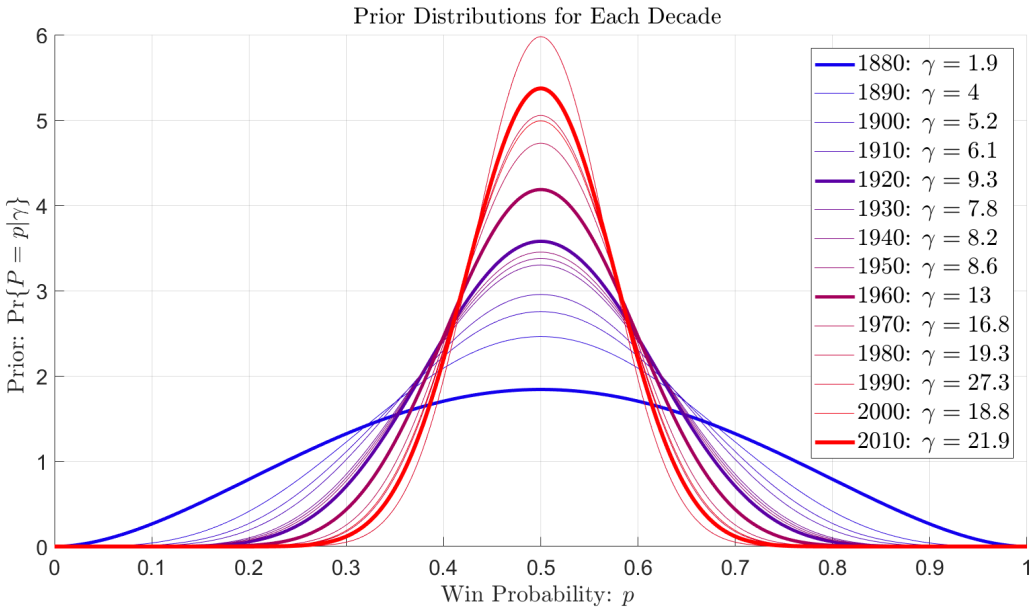

Figure 5.11: Best fit prior distributions for each decade of Major League Baseball since 1880. Decades are color coded. 1880 is shown in blue, and 2010 is shown in red. Highlighted decades are 1880, 1920, 1960, and 2010. Note the gradual concentration of the prior distribution about win probabilities equal to one half.

scores are widely available, as are player and team statistics. Considerable effort has been devoted within the sports literature to estimating win probabilities based on team/player statistics (cf. [121]). Game scores and player statistics could be leveraged to get better win probability estimates, possibly enabling the sort of network-level analysis performed for the political and animal examples described in this chapter.

An example is developed in Appendix C, where game scores are modelled as Poisson random variables. Specifically, for each pair of teams A and B it is assumed that team A has an expected scoring rate against B, and team B has an expected scoring rate against team A. Then the game is broken into a series of intervals in which teams have the opportunity to score, and the score realized after each interval is a Poisson random variables. For baseball we break the game into nine innings. If a team leads after nine innings then they win, otherwise additional innings are added, and play continues into extra innings. Play stops

once one team leads. Essentially the same model could be applied to basketball, or other sports with the same tie-breaking structure. An estimation framework for the HHD based on this Poisson scoring model is discussed in Appendix C.

We plan to test this estimation framework on baseball and basketball in the future. Similar methods could be applied to sports where a winner is determined by a sequence of repeated interactions. For example, the probability of winning in tennis can be computed from the probability of winning a single point [149], so tennis win probabilities could be estimated using scores.

5.8 Summary

In this chapter we have shown how the HHD can be used to quantify how much cyclic and transitive competition is present within a competitive system, and how the components of the HHD can be estimated from win/loss data. Our estimation tools are rooted in logistic regression, and are accompanied by uncertainty quantification, bias estimation, and hypothesis testing. Quantifying uncertainty and bias is essential since there is not always enough information in win/loss data to avoid errors. The hypothesis testing framework, and search for credible bounds, can be used to answer questions that require less information to answer than direct estimation of the components, and are less prone to errors. Examples from politics and animal behavior were presented. Future work could apply these techniques to other competitive systems.

Part IV

Dynamics: Application to Markov Processes

Chapter 6

Random Walk Models and Physical Interpretation

"The recognition of the formal analogy between the two systems of ideas leads to a knowledge of both, more profound than could be obtained by studying each system separately [237]."

– Clerk Maxwell

6.1 Preface

Thus far we have focused on using the HHD to describe the structure of an edge flow. If an edge flow describes the dynamics of a system then the HHD can also be used to analyze those dynamics. In this chapter we apply the HHD to discrete-space continuous-time Markov processes, and show that, for the appropriate choice of edge flow, using the HHD to analyze a random walk is equivalent to introducing and analyzing the thermodynamics of an analogous physical system. This analogy to thermodynamics motivates

some of the terminology (i.e. potentials), and helps to contextualize the analysis.[1] Other authors have considered similar thermodynamic interpretations of cycle decompositions of stochastic processes. Qian et al. introduce an axiomatic thermodynamic theory for diffusion processes governed by an SDE [20] which makes use of a continuous HHD, and apply it to thermal ratchets [239]. Our approach is based on Schnakenberg's network theory [5], and is closely related to Qian's theory on discrete state spaces. An extensive review of cycle decompositions of Markov chains is available in [3].

6.2 Continuous Time Discrete Space Markov Processes

Continuous-time discrete-space Markov chains are a ubiquitous class of models that are used across disciplines (cf. [75, 240, 24, 241, 242]). A continuous-time discrete-space Markov chain is a random walk on a directed network whose vertices represent possible states of the process and whose edges represent possible transitions. Let $X(t)$ denote the state of the process at time t, and x denote the a particular state. The sequence of states, $\{X_1, X_2, ...\}$ visited by the random walk is the skeleton process, and the sequence of event times $T = \{T_1, T_2, ...\}$ record the moment of each transition.

Transitions occur at random times, and occur with exponentially distributed waiting times. There is an instantaneous transition rate associated with each edge which parametrizes the corresponding waiting time distribution [241]. The transitions rates could depend on time, in which case the Markov chain is time inhomogeneous. We will only consider time homogeneous rates, that is, rates that do not change in time.

Suppose that at time t the process is in state $X(t) = x$. Then let $\mathcal{N}_x$ be the set of all states y that can be reached from x in a single transition. Let l_{yx} be the transition rate from x to y. Then, the waiting time distribution to the first transition out of state

x is exponential with parameter equal to the sum of all the transition rates $\sum_{y \in \mathcal{N}_x} l_{yx}$ so that the probability that a transition occurs after waiting a time s is proportional to $\exp\left(-\sum_{y \in \mathcal{N}_x} l_{yx}\right)$. Then, once a transition occur, the probability that the transition moved X to state y is $l_{yx} / \sum_{z \in \mathcal{N}_x} l_{zx}$ [242]. Equivalently, if a waiting time is sampled for each possible transition from exponential distributions with rates l_{yx} for each $y \in \mathcal{N}_x$ then the transition with the shortest waiting time occurs [243]. Exact trajectories of continuous-time discrete-space Markov chains can be generated by Gillespie's Direct Method, sometimes called the Stochastic Simulation Algorithm, which uses the former approach (sample the transition time, then sample which transition occured) [242], or by First Reaction Method (sample a sequence of transition times, then pick the transition which occured first). The waiting times for the unused transitions can be reused to minimize the number of random number draws when using the Next Reaction Method [243].

An important class of discrete-space continuous-time Markov processes are reaction networks. Reaction networks are widely used to model well-mixed systems of chemical reactions with small particle counts, especially in molecular biology [241]. Reaction networks are also used to simulate birth-death processes in ecology [75].

In a reaction network the set of possible transitions is described by a list of possible reactions, where each reaction makes a specific change to the state variables, and occurs with a rate that is a function of the state variables [242]. Let $X \in \mathbb{Z}^d$ represent the state of a process with d state variables. These might represent the number of molecules of a certain type, number of proteins in a certain configuration, or number of individuals of a certain species. Then let $\mathcal{R}$ be a set of possible reactions. Let $r_k \in \mathcal{R}$ be a particular reaction. Then let $s^{(k)}$ be the k^{th} stoichiometry vector, and let $\lambda_k(x) \geq 0$ be the propensity, or rate function, associated with the k^{th} reaction. Then, if at some time t the process is in state $X(t) = x$ then the set of possible reactions are all $r_k \in \mathbb{R}$ such that $\lambda_k(x) > 0$, the

instantaneous rates of the reactions are given by $\lambda_k(x)$, and, if reaction k occurs at time $t+s$ then $X(t+s) = x + s^{(k)}$ [241].

Now suppose that, at time t, the probability that $X(t) = x$ is $p(x,t)$. Then $p(x,t)$ is governed by the system of linear differential equations [241, 242]:

$$\frac{d}{dt}p(x,t) = \sum_{r_k \in \mathcal{R}} \lambda_k(x - s^{(k)})p(x - s^{(k)}, t) - \lambda_k(x)p(x,t). \tag{6.1}$$

This is the master equation for a reaction network. A master equation governs the spread of probability across the possible states of the process [242].

If the set of possible states is finite then the master equation can be expressed using a matrix L. Index the states of the process so that x_j is the j^{th} state corresponding to vertex j of $\mathcal{G}$, where $\mathcal{G}$ is the undirected version of $\mathcal{G}_{\rightleftarrows}$ which has one vertex for each possible state the system could reach, and a directed edge associated with each possible transition. Then let l_{ji} be the instantaneous rate of transition from i to j, and let $l_{ii} = -\sum_{j \in \mathcal{N}_i} l_{ji}$. Let $p_i(t)$ denote the probability that $X(t) = x_i$. Then the master equation is simply:

$$\frac{d}{dt}p(t) = Lp(t). \tag{6.2}$$

The matrix L is the Laplacian for the random walk. Note that this is not the same as the node or face Laplacians that appeared in the HHD. The columns of L all sum to zero, so probability is conserved by Equation (6.2).

Given the transition rates l_{ij} we will study the edge flow:

$$f_k = \frac{1}{2}\left(\log(l_{j(k)i(k)}) - \log(l_{i(k)j(k)})\right) \tag{6.3}$$

Note that the edge flow between i and j is only finite if it is possible to move from i to

j, and from j to i, with nonzero probability. Thus we assume that, if there is a directed edge from i to j with nonzero transition rate, then there must be a directed edge from j to i with a nonzero transition rate. This assumption is based on the principle of microscopic reversibility, which holds for any physical system. That said, not all models of physical systems obey microscopic reversibility since it may be convenient to round very small transition rates to zero, especially if they are too small to measure. This limits the applications of our methods to continuous-time discrete-space Markov chains without absorbing states [75]. Many ecological models include absorbing states, namely, extinction, so it is important to note this limitation.

The motivation for this choice of edge flow is described in Section 6.3, where it is shown that this choice of edge flow arises naturally from a class of networks whose dynamics do not exhibit any tendency to cycle, and can be described entirely by a potential function and a set of conductances on the undirected edges. In Section 6.5 we show that this choice of edge flow is naturally related to the thermodynamics of Markov chain models of physical processes. Thus, by defining the edge flow according to Equation (6.3) the analysis using the HHD amounts to an analysis of the thermodynamics of an analogous physical system with the same transition rates as the system of interest.

Under the assumption of microscopic reversibility the directed graph $\mathcal{G}_{\rightleftarrows}$ has a pair of edges between each pair of nodes that are connected, so one undirected edge can be introduced for each pair of directed edges. As in Chapter 2 let $\mathcal{G}$ denote the undirected version of $\mathcal{G}_{\rightleftarrows}$. If $\mathcal{G}$ is connected then $\mathcal{G}_{\rightleftarrows}$ is irreducible. A Markov chain is irreducible if there is a path between any pair of nodes that can traversed with nonzero probability [241]. Since $\mathcal{G}$ is connected there is a path from any i to any j in $\mathcal{G}$. Microscopic reversibility ensures that the path can be taken in either direction in $\mathcal{G}_{\rightleftarrows}$ with nonzero probability. If a continuous-time discrete-space Markov chain is irreducible then it is ergodic, and $p(t)$

converges to a unique steady state distribution, q, from any initial conditions [241, 244]. This steady state is the Perron-Frobenius eigenvector of the matrix L corresponding to eigenvalue 0, and is the unique vector in the nullspace of L with nonnegative entries that all sum to one.

The existence of a unique steady state, and convergence to it, can be proved directly from the Perron-Frobenius Theorem and the irreducibility of $\mathcal{G}_{\rightleftarrows}$. Suppose that $p(0) = p_0$. Then since $p(t)$ obeys the master equation (Equation (6.2)) the probability distribution at time t is $\exp(Lt)p_0$ where $\exp(A)$ denotes the matrix exponential. Then, if we coarse grain time by considering a sequence of discrete times $\{0, t_1, t_2, ...\}$ where $t_{j+1} = t_j + \Delta t$, then $X(t_j)$ is a discrete time Markov process with transition matrix $\exp(L\Delta t)$. Since $\exp(L\Delta t)$ is a transition matrix for a discrete time Markov process it must be a stochastic matrix, that is, a matrix with all nonnegative entries and whose columns sum to one to conserve probability. Therefore $\exp(L\Delta t)$ is a square matrix with real, nonnegative entries. Moreover, since $\mathcal{G}_{\rightleftarrows}$ is irreducible, and any sequence of nonzero waiting times is possible, it is possible that if $X(t) = x$ then $X(t+\Delta t) = y$ for any $y \in \mathcal{V}$ provided $\Delta t > 0$ [245]. Then all of the entries of $\exp(L\Delta t)$ must be positive if $\Delta t > 0$, so the discrete time process is both irreducible and aperiodic [244]. The Perron-Frobenius Theorem states that if a matrix A is square, real, and non-negative then it has a unique largest eigenvalue corresponding to a nonnegative eigenvector [246]. Any stochastic matrix has an eigenvalue equal to one since the sum of the columns equal one. The Gershgorin circle theorem[2] ensures that this is the largest possible eiegenvector since all entries are nonnegative and all columns sum to one. It follows that there is a unique eigenvector corresponding to eigenvalue 1, and all of

[2] All eigenvalues of a square matrix A with complex entries are contained in the union of the disks centered at the diagonal entries of the matrix, and with radii equal to the sum of the magnitude of each off-diagonal component in the corresponding columns [247]. For a real $V \times V$ matrix A it follows that all eigenvalues are contained in the interval $\cup_{j=1}^{V}[a_{jj} - \sum_{i\neq j}|a_{ij}|, a_{jj} + \sum_{i\neq j}|a_{ij}|]$. For a stochastic matrix $a_{jj} + \sum_{i\neq j}|a_{ij}| = \sum_i a_{ij} = 1$ and $a_{jj} - \sum_{i\neq j}|a_{ij}| = 0$.

its entries have the same sign, so it can be chosen so that all of the entries are nonnegative. Properly normalized this eigenvector is a steady state distribution. All other eigenvalues must be less than one, and, by Gershgorin, greater than or equal to zero. Thus, no matter the initial distribution, the sequence $p(t_j)$ must converge to the steady state distribution [241].

Lemma 21 (Convergence to a Unique Steady State)**.** *If $\mathcal{G}_{\rightleftarrows}$ is a finite connected directed network that obeys microscopic reversibility (if there is an edge from i to j with nonzero transition rate then there is an edge from j to i with nonzero transition rate), then the continuous-time discrete-space Markov process $X(t)$ with Laplacian L has a unique steady state distribution q, and the probability that $X(t) = x$, $p(x,t)$, converges to q from any initial distribution $p(x,0) = p_0(x)$.*

6.3 The Conservative Case: Detailed Balance

When we studied tournaments we started by looking for a special class of tournaments that were, in a sense, acyclic. Here, as before, we start by considering a special case - the conservative case. We would like to choose an edge flow so that, when the edge flow is conservative, the process has no tendency to circulate.

One way to think about circulation is through the probability fluxes, $J(t)$. The probability flux on edge k is the difference between the rate of flow of probability from $i(k)$ to $j(k)$, and the rate of flow from $j(k)$ to $i(k)$:

$$J(t)_k = J(t)_{i(k)j(k)} = l_{j(k)i(k)} p(t)_i(k) - l_{i(k)j(k)} p(t)_j(k). \tag{6.4}$$

Note that the collection of fluxes is, itself, an edge flow.

A reasonable measure of circulation is the curl of the fluxes. If the curl of the flux around a loop is nonzero then the process has a net tendency to move around that loop in a particular direction. The fluxes do not circulate if the curl of the fluxes around any loop is zero, or, more strongly, if it is impossible to follow the fluxes around a cycle. More precisely, the fluxes do not circulate if there is no cycle where all of the fluxes on the cycle point in the same direction. The fluxes depend on the distribution, so different p will lead to different J. It would be easier to define a "no circulating flux" condition if the fluxes were unique. Since the probabilities $p(t)$ converge to a unique steady state q, the fluxes converge to a unique set of steady state fluxes. Therefore, one way to define a process that does not tend to circulate is a process whose steady state fluxes do not circulate. We will show in Section 7.3.2 that the requirement that the steady state fluxes do not circulate actually implies that the fluxes never circulate, regardless the current distribution.

The rate of change in the probability at a particular node is the (negative) divergence of the fluxes:

$$\frac{d}{dt}p_i(t) = [G^\intercal J(t)]_i \tag{6.5}$$

At steady state $\frac{d}{dt}p(t) = 0$ so the steady state fluxes must be divergence-free. By Theorem 5, if $\mathcal{G}$ is finite and closed, then the steady state fluxes must lie in the range of $C^\intercal$. Therefore:

Lemma 22. *The curl of the steady state fluxes around any loop is zero if and only if there is no cycle on which all of the steady state fluxes point in the same direction, and if the curl of the steady state fluxes is zero then the steady state fluxes are all identically zero.*

Proof. First, suppose that all of the steady state fluxes are zero. Then $CJ = C0 = 0$ so the curl of the fluxes is zero. If $J = 0$ then the fluxes do not point in any direction, so it is impossible to find a cycle on which all of the fluxes point in the same direction (clockwise

or counterclockwise).

Next, suppose that there is no cycle in $\mathcal{G}$ on which all of the fluxes point in the same direction (clockwise or counterclockwise). Form a directed graph, $\mathcal{G}_{\rightarrow}$, with one edge for each undirected edge of $\mathcal{G}$ with a nonzero flux, oriented in the direction of the flux. The steady state fluxes are divergence free so the directed graph, $\mathcal{G}_{\rightarrow}$ must be cyclic (see Lemma 16). That is, for any pair of connected nodes in $\mathcal{G}_{\rightarrow}$ there is a cycle including both nodes. Therefore, if there are any connected nodes in $\mathcal{G}_{\rightarrow}$ there is a cycle in the original graph that can be traversed entirely in the direction of the flux. It follows that, if there is no such cycle in the original graph, then there must be no directed edges in $\mathcal{G}_{\rightarrow}$, so all of the fluxes must be zero. If $J = 0$ then the curl of J is zero on any cycle.

Finally, suppose that the curl of the steady state fluxes is zero. The steady state fluxes are divergence free so there exists a vector $\theta_J \in \mathbb{R}^L$ such that $C^\intercal \theta_J = f_{\text{rot}}$. Then, if the curl of J is zero on any loop, $CC^\intercal \theta_J = C^\intercal f_{\text{rot}} = 0$. The matrix $CC^\intercal$ is invertible so $CC^\intercal \theta_J = 0$ implies $\theta_J = 0$, and hence $J = 0$.

□

Lemma 22 can be shown more intuitively as follows. Suppose the system is at steady state and there is a nonzero net flow of probability between two states. In order for the two states to maintain constant probability the net flow into each must balance the net flow out of each. Since there is a net flow between the two there must be an equivalent net flow into and out of the pair. If this is true for all states then any net flow of probability must form closed cycles (see Lemma 16). Therefore, if probability never flows in a closed cycle, at equilibrium the net flow of probability between any two states must be zero.

Lemma 22 establishes the equivalence of two different notions of circulation. Whether we define circulation as a nonzero curl, or the existence of a cycle on which all of the fluxes point the same direction, the only steady state in which the fluxes do not circulate

is a steady state with no flux. This requirement is intuitively connected to conservative dynamics. If the probability flow is driven by a scalar potential then it cannot circulate when at steady state. Probability cannot flow downhill around a closed cycle, since no path around a closed cycle can be downhill at every step.

The requirement that all of the steady state fluxes are zero is called *detailed balance* [248]. Isolated physical systems automatically reach an steady state without net circulation since, in physical systems, rotation at steady state requires an external energy source [240]. Thus all isolated physical systems obey detailed balance. A steady state with zero steady state fluxes is called an equilibrium. If the process does not obey detailed balance it will still reach a steady state, but the steady state is a nonequilibrium steady state since the probability fluxes are nonzero and circulate [248].

A closely related idea is time-reversibility [20]. A Markov process is reversible if any trajectory is equally likely forward or backward conditioned on the endpoints. A process that tends to circulate is not time-reversible since trajectories moving around cycles will usually traverse the cycles in a particular direction in forward time, and the opposite direction in backward time. Thus if a process is time-reversible it should not circulate. This approach offers a more general notion of non-circulation since it is not defined with respect to steady state dynamics. That said, the steady state fluxes equal the average rate at which an edge is crossed in its forward direction minus the average rate it is crossed in its backward direction, so can be expressed as the average number of forward traversals minus the average number of backward traversals on a long trajectory [4, 6]. If the process is time-reversible then the probability of traversing any path forwards should equal the probability of traversing the path backwards, so the net fluxes should approach zero on long trajectories. This suggests that the two notions may actually be equivalent, even though time-reversibility is defined without referencing a steady state distribution.

Consider a finite trajectory visiting the sequence of states $\{x_0, x_1, ...x_n\}$ and spending waiting times $w = \{w_0, w_1, ...w_n\}$ in each state. Then define the backward trajectory: $\{x_n, ...x_1, x_0\}$, $\{w_n, ...w_1, w_0\}$. Since the random walk is a Markov process, the probability of each transition and event time depends only on the current state. Therefore, if we let X denote the states visited by the process $X(t)$, and let W denote the waiting times, then the probability of observing the forward trajectory out of all trajectories of length n starting at x_0 is:

$$\begin{aligned}
&\Pr\{X = \{x_0, x_1, ...x_n\}, W = \{w_0, w_1, ...w_n\}\} = \\
&\qquad = \left[\prod_{k=0}^{n-1} \frac{l_{x_{k+1}x_k}}{|l_{x_k x_k}|} |l_{x_k x_k}| \exp(-|l_{x_k x_k}| w_k)\right] |l_{x_n x_n}| \exp(-|l_{x_n x_n}| w_n) \\
&\qquad = \left[\prod_{k=0}^{n-1} l_{x_{k+1}x_k} \exp(-|l_{x_k x_k}| w_k)\right] |l_{x_n x_n}| \exp(-|l_{x_n x_n}| w_n) \\
&\qquad = \left[\prod_{k=0}^{n-1} l_{x_{k+1}x_k}\right] \left[\prod_{k=0}^{n} \exp(-|l_{x_k x_k}| w_k)\right] |l_{x_n x_n}|
\end{aligned} \tag{6.6}$$

where $|l_{x_k x_k}| = \sum_{y \in \mathcal{N}_{x_k}} l_{y x_k}$ is the net rate of transition out of state x_k. Then the first bracketed expression depends on the direction of traversal, the middle bracketed expression does not depend on the direction of traversal since it only depends on the states visited and time spent in each state, and the last term depends on the time spent in the last node. Therefore the ratio of the probability of the forward to backward trajectories is:

$$\left[\prod_{k=0}^{n-1} \frac{l_{x_{k+1}x_k}}{l_{x_k x_{k+1}}}\right] \frac{|l_{x_n x_n}|}{|l_{x_0 x_0}|}$$

The first half of this expression depends on the path taken, and the second half depends on the conditioning at the endpoints. For example, if we start recording the trajectory as soon as it leaves the first state and stop recording as soon as it arrives at the second state,

then the ratio of the forward to backward trajectories has the same form. Alternatively, if we only consider the probabilities of the skeleton process then we arrive at the same expression.[3]

Suppose that the trajectory is a cycle. Then $\frac{|l_{x_n x_n}|}{|l_{x_0 x_0}|} = 1$ since $x_0 = x_n$. Therefore the ratio of the probability of observing a cycle in one direction to the probability of observing the cycle in its reverse direction is:

$$\left[\prod_{k=0}^{n-1} \frac{l_{x_{k+1}x_k}}{l_{x_k x_{k+1}}}\right].$$

A Markov chain is reversible if these two probabilities are equal [240]. That is, a Markov chain is reversible if the probability of observing a cyclic sequence of events does not depend on the direction in which the cycle is traversed. If these two probabilities are the same for all cycles then the process obeys the cycle condition (see Equation (4.7)):

$$\prod_{k=0}^{n-1} \frac{l_{x_{k+1}x_k}}{l_{x_k x_{k+1}}} = 1 \text{ if } x_n = x_0. \tag{6.7}$$

Lemma 23. *If a discrete-space continuous-time Markov process $X(t)$ on a finite network with microscopic reversibility obeys detailed balance (steady state fluxes $J = 0$), then it is reversible and satisfies the cycle condition.*

Proof. Suppose that the process obeys detailed balance. Then, by definition, all of the steady state fluxes are zero so:

$$l_{ji} q_i = l_{ij} q_j. \tag{6.8}$$

for all connected pairs of nodes i and j. Now consider a closed cycle y starting and ending

[3]For the skeleton process the probabilities are $\prod_{k=0}^{n-1} l_{x_{k+1}x_k}/|l_{x_k x_k}|$ and $\prod_{k=n}^{1} l_{x_{k-1}x_k}/|l_{x_k x_k}|$ so the ratio of forward to backward probabilities is the same.

in state x_i. For clarity consider the smallest nontrivial cycle:

$$y = x_i \to x_j \to x_k \to x_i.$$

By rearranging the detailed balance condition we can fix q_j in terms of q_i, q_k in terms of q_j, and q_i in terms of q_k:

$$q_j = \frac{l_{ji}}{l_{ij}} q_i, \quad q_k = \frac{l_{kj}}{l_{jk}} q_j, \quad q_i = \frac{l_{ki}}{l_{ik}} q_k.$$

Carrying this process to completion solves for q_i in terms of q_i. That is, carrying the process to completion gives a consistency condition for detailed balance that is entirely independent of the steady state distribution. Plugging the first equation into the second, and the second into the third:

$$q_i = \frac{l_{ji}}{l_{ij}} \frac{l_{kj}}{l_{jk}} \frac{l_{ik}}{l_{ki}} q_i$$

Or:

$$\frac{l_{ji}}{l_{ij}} \frac{l_{kj}}{l_{jk}} \frac{l_{ik}}{l_{ki}} = 1, \quad l_{ij} l_{jk} l_{ki} = l_{ik} l_{kj} l_{ji}. \tag{6.9}$$

Now consider a cyclic sequence of nodes, $x_0, x_1, x_2, ...x_n = x_0$. Then:

$$q_{x_k} = \left[\prod_{j=0}^{k-1} \frac{l_{x_{j+1} x_j}}{l_{x_j x_{j+1}}} \right] q_{x_0}.$$

Since the path is a cycle $q_{x_n} = q_{x_0}$ so the product of the ratio of forward to backward rates around the cycle must equal 1. It follows that, if a process obeys detailed balance then it must obey the cycle condition (Equation (6.7)). If a process obeys the cycle condition then it is time reversible, so any process that obeys detailed balance (zero steady state fluxes) must be reversible.

□

Lemma 23 shows that any process that obeys detailed balance is also reversible. The converse is also true [20]. If a process is reversible then it satisfies the cycle condition, and the cycle condition implies that the process has zero steady state flux.

Theorem 24 (Detailed Balance)**.** *A discrete-space continuous-time Markov chain on a finite network obeying microscopic reversibility obeys detailed balance if and only if:*

1. *the steady state fluxes are all zero,*
2. *it is time-reversible (satisfies the cycle condition Equation (4.7)),*
3. *the edge flow* $f_k = \frac{1}{2}\left(\log(l_{j(k)i(k)}) - \log(l_{i(k)j(k)})\right)$ *is conservative,*
4. *there exists a potential function* ϕ *such that* $-G\phi = f$ *and* $l_{ij} = \rho_{ij}\exp(\phi_j - \phi_i)$ *where* $\rho_{ij} = \rho_{ji}$ *is a symmetric function on the edges*

If a process obeys detailed balance then the steady state distribution obeys a Boltzmann type distribution:

$$q_i \propto \exp(-2\phi_i). \tag{6.10}$$

Proof. Lemma 23 establishes that detailed balance implies the cycle condition, and, as a consequence, reversibility. If the cycle condition holds then the product of the ratio of forward to backward transition rates around any cycle equals one. Therefore, if $x_0, x_1, ..., x_n$ is a cyclic sequence of nodes:

$$\log\left(\prod_{j=0}^{n-1}\frac{l_{x_{j+1}x_j}}{l_{x_jx_{j+1}}}\right) = \sum_{j=0}^{n-1}\log(l_{x_{j+1}x_j}) - \log(l_{x_jx_{j+1}}) = \log(1) = 0.$$

Thus, if the edge flow f_k is defined according to Equation (6.3), then the cycle condition

implies that the curl of f around any loop is zero. It follows from the fact that the network is finite that f must be conservative. If f is conservative then there must exist a potential ϕ such that $-G\phi = f$. If $f = -G\phi$ then $\sqrt{l_{ij}/l_{ji}} = \exp(-(\phi_i - \phi_j))$.

Then:

$$l_{ij} = \sqrt{l_{ij}l_{ji}}\sqrt{\frac{l_{ij}}{l_{ji}}} = \rho_{ij}\exp(\phi_j - \phi_i)$$

where $\rho_{ij} = \sqrt{l_{ij}l_{ji}}$ is the geometric average of the forward and backward transition rates, so does not depend on the ordering of the indices. Note that these coefficients have the same units as the transition rates, probability over time, while the geometric difference used to define the edge flow and associated potential is unit-less.

To prove the converse suppose that $f = -G\phi$ for some potential. Then the definition of the edge flow implies $l_{ij} = \rho_{ij}\exp(\phi_j - \phi_i)$. If $f = -G\phi$ it is automatically conservative, so is curl free. Since the edge flow is conservative the sum of f around any loop is zero, so the cycle condition is automatically satisfied and the process is reversible. It remains to show that, if the cycle condition is satisfied, then the steady state fluxes are all zero.

Any reversible physical process is energetically isolated, so reaches thermal equilibrium, with a steady state fixed by the Boltzmann distribution. Therefore, a natural ansatz for the steady state of a process obeying detailed balance is:

$$q_i = \frac{1}{Z}\exp(-\beta\phi_i). \tag{6.11}$$

where $Z = \sum_i \exp(-\beta\phi_i)$ is analogous to the partition function and β is analogous to the inverse temperature. To check if there is a choice of β for which Equation (6.11) defines the steady state distribution compute the flux on each edge:

$$l_{ji}q_i - l_{ij}q_j = \rho_{ij}\left(\exp(\phi_i - \phi_j)\exp(-\beta\phi_i) - \exp(\phi_j - \phi_i)\exp(-\beta\phi_j)\right).$$

Now, if $\beta = 2$ then:

$$l_{ji}q_i - l_{ij}q_j = \rho_{ij}\left(\exp(-(\phi_i + \phi_j)) - \exp(-(\phi_i + \phi_j))\right) = 0$$

so the flux on each edge is zero. If the flux on every edge is zero then the rate of change of probability at each node is zero, so q defined by Equation (6.11) is the steady state. Since the steady state fluxes are all zero the process obeys detailed balance.

Alternatively, detailed balance requires $l_{ij}q_j = l_{ji}q_i$ or:

$$\frac{q_i}{q_j} = \frac{l_{ij}}{l_{ji}}. \tag{6.12}$$

But then $\log(q_i) - \log(q_j) = 2f_{ij}$. Since $f = -G\phi$ this implies $\log(q_i) - \log(q_j) = 2(\phi_j - \phi_i)$ which is solved, up to the addition of a constant, by setting $\log(q_i) = -2\phi_i$. □

Therefore, for networks obeying detailed balance we have a simple decomposition of the edge transition rates that expresses the edge transition rates in terms of a potential:

1. Split each pair of transition rates into their geometric difference $\sqrt{l_{ij}/l_{ji}}$ and geometric average $\sqrt{l_{ij}l_{ji}}$. Let $\rho_{ij} = \sqrt{l_{ij}l_{ji}} = \rho_{ji}$ denote the geometric average of the forward and backward rates. We will refer to ρ as the per capita conductances, since ρ have units one over time, and correspond to the average transition rate over the edge. This choice of terminology is motivated in Section 6.3.1. Then set the edge flow f equal to the log of the geometric difference: $f_{ij} = \log(\sqrt{l_{ij}/l_{ji}})$.

2. If the process obeys detailed balance then $f = -G\phi$ where ϕ is the scalar potential, and the steady state $q_i \propto \exp(-2\phi)$. Note that the steady state in detailed balance is independent of the per capita conductances ρ. Whenever it is possible to express

f with the gradient of a potential the corresponding Markov process obeys detailed balance and is time reversible.

In contrast, if f is not in the range of the gradient then it is not conservative and the system does not obey detailed balance. Outside of detailed balance the steady state is maintained by circular balance [4], with nonzero steady state currents. In that case the steady state distribution does not obey Equation (6.10). By analogy the *effective potential* is defined:

$$\phi_{eff} \propto -\log(q) \tag{6.13}$$

for an appropriately chosen scaling constant.

6.3.1 An Electric Circuit Analogy

When the Markov process obeys detailed balance we can introduce an electric circuit analogy where the flow of current over the circuit mimics the flow of probability over the network. This analogy is used as an introductory example for how a random walk on a network can be related to a physical process, and how that relation can help build intuition about the random walk. The usefulness of any analogy of this kind depends on the familiarity of the analogous system, and how contrived the analogy is. For a random walk obeying detailed balance there is a simple analogy to circuits that clarifies the role of the potential, and the geometric average of the transition rates which appeared in the previous discussion.

Connections between electrical networks and random walks on graphs are well studied (cf. [28, 249]). Given a network an associated electrical network is constructed by replacing each edge in the network with a resistor. By scaling the resistances appropriately the electrical network can imitate the behavior of the random network. Without an energy source

(battery) to drive current the electric circuit analogy only applies to reversible random walks (conservative networks). In this context charge, current, and voltage all have a probabilistic interpretation [28]. Given different boundary conditions (input current or voltages) we can ask different questions about the random network. For example, the voltage established when a current is introduced at some nodes, and removed at a boundary set, is analogous to passage times onto the boundary set in the random network. Commute and cover times in the random network are equivalent to the total resistance between a pair of nodes [249].

These connections are both illuminating and useful. They can be used to help clarify dynamics on random network processes, and to guide intuition. Since efficient methods for computing current, charge, and voltage distributions are well developed, the connection between random network processes and electrical networks can be leveraged to solve large problems in random networks. For this reason circuit theory has been used to study connectivity in chemical, neural, economic, and social networks [250]. Circuit theory has also been introduced in ecological settings to study connectivity in spatial dispersal networks [250].

Suppose you are given an electric circuit that consists exclusively of resistors joining nodes. The nodes are analogous to the states of the network, the resistors are analogous to the undirected edges. Each node is attached to a capacitor which is subsequently grounded as shown in Figure 6.1. When we introduce a distribution of charge Q over the nodes, how will the charge flow through the network?

Suppose the capacitance of node x_i is C_i. Then the voltage of node x_i is $V_i = Q_i/C_i$. The capacitance is the capacity of a node to store charge at a given voltage. If node x_i is connected to node x_j with a resistor of resistance R_{ij} then the current I_{ij} across the resistor from x_i to x_j is:

$$I_{ij} = \frac{1}{R_{ij}} \left(\frac{Q_i}{C_i} - \frac{Q_j}{C_j} \right).$$

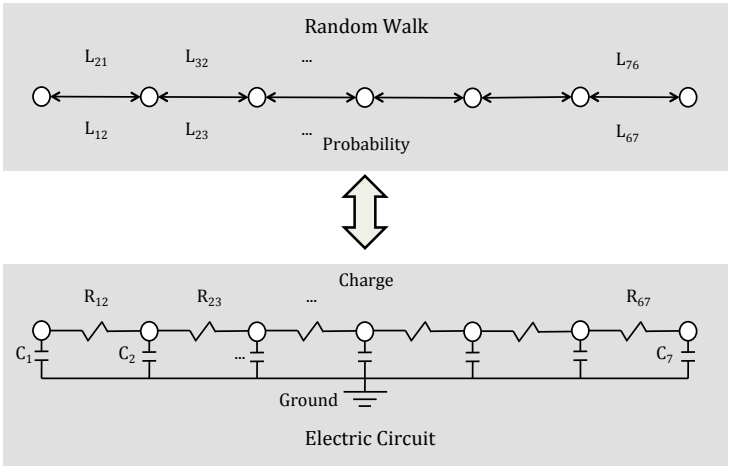

Figure 6.1: The electric circuit analogy. Charge is analogous to probability, and will tend to accumulate where the capacitance is large. A large capacity to store probability corresponds to small potentials. The resistances scale the time for charge to distribute itself, but have no effect on the final distribution.

The flow of current is remarkably similar to the rate of flow across an edge in our original network:

$$l_{ji}p_i - l_{ij}p_j = \rho_{ij}\left(\exp\left(-[G\phi]_{ji}\right)p_i - \exp\left(-[G\phi]_{ij}\right)p_j\right).$$

however there are subtle differences between the two equations. If we consider charge analogous to probability then Q is analogous to p. Note that for the electric network the charge is divided by a function on the nodes C_i, whereas for the probabilistic network the distribution is multiplied by a function on the edges $\exp\left(-[G\phi]_{ij}\right)$. In order to make these two equations consistent we need to find a way to rewrite the second equation as a difference between functions on the nodes.

The equations also differ in units. The electrical resistance R has unit volt seconds per

coulomb: $[R] = [V][t]/[Q]$. The capacitance C has unit coulomb per volt: $[C] = [Q]/[V]$. Since $\exp(-[G\phi]_{ij})$ is a geometric difference of two functions with the same units it is necessarily dimensionless. The per capita conductance ρ is the geometric mean of the instantaneous rates, so necessarily has units time: $[\rho] = 1/[t]$.

The differences here are subtle but important. Suppose we measured probability in Coulombs. Treating energy as a dimensionless quantity (ignoring Volts) $[\rho][Q] = 1/[R]$ and $[Q]/[C]$ equals $[\exp(-(\nabla\phi)_{ij})]$. That is, the latter are per capita rates, the former are not. We can treat energy as dimensionless since the energies are related to the edge flow f, which is dimensionless.

To resolve these differences expand the flow of probability:

$$l_{ji}p_i - l_{ij}p_j = \sqrt{l_{ij}l_{ji}}\left(\sqrt{\frac{l_{ji}}{l_{ij}}}p_i - \sqrt{\frac{l_{ij}}{l_{ji}}}p_j\right) = \sqrt{l_{ij}l_{ji}}\left(\sqrt{\frac{q_j}{q_i}}p_i - \sqrt{\frac{q_i}{q_j}}p_j\right).$$

where the last equality follows from Equation (6.12). Next, simplify the ratio of the steady state distributions and pull out a common factor of the geometric averages of the steady states:

$$\sqrt{l_{ij}l_{ji}}\left(\sqrt{\frac{q_j}{q_i}}p_i - \sqrt{\frac{q_i}{q_j}}p_j\right) = \sqrt{l_{ij}l_{ji}}\left(\frac{\sqrt{q_iq_j}}{q_i}p_i - \frac{\sqrt{q_iq_j}}{q_j}p_j\right) = \sqrt{l_{ij}l_{ji}q_iq_j}\left(\frac{p_i}{q_i} - \frac{p_j}{q_j}\right).$$

Notice that this form matches the electrical network equation if we let $C = q$ and $1/R_{ij} = \rho_{ij}\sqrt{q_iq_j}$. Then the per capita conductance ρ is replaced with the resistance R and the capacitance ratio with the capacitance C. So, letting:

$$\frac{1}{R_{ij}} = \sqrt{l_{ij}l_{ji}q_iq_j} = \rho_{ij}\sqrt{q_iq_j}, \quad C_i = q_i = \exp(-2\phi_i) \tag{6.14}$$

gives an exact analogy between the flow of probability on the probabilistic network and the

flow of charge on a electric network:

$$I_{ij} = l_{ji}p_i - l_{ij}p_j = \frac{1}{R_{ij}}\left(\frac{p_i}{C_i} - \frac{p_j}{C_j}\right). \tag{6.15}$$

Notice that the conductance of the electric circuit is the per capita conductance times the geometric average of the steady state charge (probability) at each end of each edge. Therefore $1/R_{ij} = \rho_{ij}\sqrt{q_i q_j} = \sqrt{(l_{ij}q_j)(l_{ji}q_i)}$ is the geometric average of the rate at which probability is exchanged between the two nodes at steady state. Since the system obeys detailed balance these two rates are identical, so the conductance $1/R_{ij}$ is the rate at which charge/probability flows between the two nodes at steady state.

This analogy provides a clear interpretation of ρ and ϕ. In particular it shows that we can think of ϕ as the logarithm of a capacitance associated with each node. This should not be a surprise, after all, capacitance is a per capita potential. Glancing back at Equation (6.15) we can also see that the capacitance is effectively the amount of probability that any node is capable of holding before it starts to diffuse out. In other words, the capacitance describes each node's capacity to store probability.

More generally this analogy shows that the diffusion of probability obeys the same physical principles as the diffusion of charge. Both represent an ensemble of random walks directed by a potential function. Analogies of this type are well explored (see [34, 249, 28, 251]), and have been proposed in an ecological setting ([250]). In general these analogies only use resistors. The analogy developed here generalizes this idea by introducing an underlying capacitance. Introducing capacitors changes the interpretation of the electric network, and puts more emphasis on the potential/equilibrium distribution.

It is not hard to imagine other physical analogies for the flow of probability.

Consider a series of tanks of water separated by pipes of different gauge. Introduce a

solute into one of the tanks. The rate at which it will diffuse between all the tanks depends on their volume, and the gauge of pipe linking the tanks. The total quantity of solute is conserved as it diffuses, and it diffuses to a static equilibrium. Solute diffuses from high concentration to low concentration. Equilibrium is achieved when all the tanks have the same concentration. Since the amount of solute needed to reach equilibrium concentration per tank depends on the volume of each tank the tank volume plays the same role as the capacitance. The concentration plays the role of voltage.

The main goal of these analogies is to provide physical intuition for network potentials. They offer natural ways to think about the role of potentials in networks, and emphasize crucial properties of conservative networks. They all move from high to low potential. They are all ergodic. They all approach a static steady state that is independent of the rate of diffusion across any edge.

We should not be surprised that evocative analogies are easy to find. One of the main virtues of a potential description is its appeal to intuition. This is why, at least in part, that qualitative potentials are frequently invoked in lieu of a quantitative potential. References to an evolutionary potential landscape, or basin of attraction often invoke potentials to make informal or verbal arguments based on that intuition (cf. [252]).

The electric circuit analogy provides a concrete example that can be used to ground intuition in a qualitative setting. In Section 6.5 we develop a much deeper physical analogy that relates the dynamics of the Markov process to the thermodynamics of an analogous physical process.

6.4 The HHD for Markov Chains

Let L be the Laplacian for a continuous-time discrete-space Markov chain, where l_{ij} is the rate of transition from node j to node i with the property that if $l_{ij} > 0$ then $l_{ji} > 0$. Then let:

$$\begin{aligned} \rho_{ij} &= \sqrt{l_{ij} l_{ji}} \\ f_{ij} &= \frac{1}{2} \left(\log(l_{ij}) - \log(l_{ji}) \right) = \log \left(\sqrt{\frac{l_{ij}}{l_{ji}}} \right) \end{aligned} \tag{6.16}$$

where f is the edge flow, and ρ are the per capita conductivites. Then $l_{ij} = \rho_{ij} \exp(f_{ij})$, $\rho_{ij} = \rho_{ji}$, and $f_{ij} = -f_{ji}$. The flow, f, represents the difference in the forward and backward rates between a pair of states, and thus the preferred reaction direction. The conductivity, ρ, describes the rate at which reactions occur, on average, between the pair.

Let $\mathcal{G}$ be the undirected version of $\mathcal{G}_{\rightleftarrows}$. Let G be the gradient operator associated with $\mathcal{G}$ and let C be the curl associated with a cycle basis of $\mathcal{G}$. Then let ϕ, θ be the solutions to:

$$f = f_{\text{con}} + f_{\text{rot}} = -G\phi + C^{\intercal}\theta \tag{6.17}$$

where θ is unique and ϕ is determined uniquely up to an additive constant.

Equation (6.17) defines the HHD for continuous-time discrete-space Markov chains. The rest of this dissertation is devoted to studying how this decomposition can be used to understand the dynamics of the Markov process.

Note that a process obeys detailed balance if and only if $f_{\text{rot}} = 0$ and $\theta = 0$. Otherwise the process does not obey detailed balance and reaches a nonequilibrium steady state, in which the probability distribution is constant, but there are nonzero circulating probability fluxes.

6.5 A Thermodynamic Analogy

In Section 6.3.1 we developed a physical analogy relating Markov processes that obey detailed balance to an electric circuit built of resistors and capacitors. This analogy is limited in two ways. First, it only treats the detailed balance case, in which the edge flow defined by Equation (6.3) is conservative. Second, the analogy relates the flow of probability over the network to the flow of charge over a circuit, which is treated as a deterministic process. A better physical analogy would relate the given Markov process to a Markov process representing a physical system with meaningful thermodynamics. Then the thermodynamics of the analogous system could be analyzed, and related to the analysis using the HHD.

The following section develops this thermodynamic description. Highlights include simple relations between the network potentials and the Free Energy of the system, demonstration that the rotational potential is associated with coupling to external energy sources, and a decomposition of the affinities (generalized thermodynamic forces) into an internal and external component via the HHD. These results are largely based on [5, 19, 20, 4] who introduce axiomatic thermodynamics for Markov chains. Although the following section only addresses the thermodynamics for discrete state spaces some of the same results carry over naturally to continuous state spaces [20]. Here we will not address continuous state spaces since we do not yet have the analytic tools to apply the HHD to diffusive processes in the continuum.

It is important to note that, if the study system is physical, then the thermodynamic interpretation developed here may not match the thermodynamics of the original system since our approach assumes no knowledge outside of the states and transition rates. Systems with different underlying thermodynamics may have the same transition rates, so at

best we hope to construct an analogous physical system with meaningful thermodynamics. Thus our objectives are primarily phenomenological. For appropriate physical systems this analysis is an exact analysis of the actual thermodynamics of the system (cf. [22]).

Assume that $X(t)$ is a Markov process on a network where x_j is the collection of state variables associated with the j^{th} possible state of the system. These are usually counts, such as the number of molecules of a certain type, in a certain configuration, or in a certain place [241]. In an ecological setting these might be the number of individuals in a certain population, in a certain region, with a certain age, and a certain sex [75]. Assume that the state space is finite, though potentially very large. Assume that the underlying directed graph $\mathcal{G}_{\rightleftarrows}$ is connected, and that for every forward transition there exists a backward transition. Let L be the Laplacian storing the transition rates l_{ij}.

The key physical principle which bridges our formal development of the HHD for Markov processes, defined in Section 6.4, and the thermodynamics of an analogous physical system is the following relation between the ratio of forward and backward transition rates and the work required to move forward over a transition [4]. If there is a physical process which is modeled by a Markov chain on $\mathcal{G}_{\rightleftarrows}$ with transition rates L, then:

$$\frac{l_{ji}}{l_{ij}} = \exp\left(\frac{1}{k_B T} w_{ij}\right) \tag{6.18}$$

where k_B is the Boltzmann constant, T is the temperature, and w_{ij} is the work required to move from state i to state j [4, 5]. Schnakenberg provides an example based on the diffusion of charged particle through a pore in a membrane [5], and Esposito et al. provide an example based on a thermal engine [253].

Taking a logarithm on both sides and dividing by 2 yields:

$$f_{ij} = \frac{1}{2k_BT} w_{ij} \tag{6.19}$$

where the factor of $1/2$ arose from the choice to define the edge flow as the log of the geometric difference in the transition rates.

Therefore the edge flow defined by Equation (6.3) is analogous to the work required to move between two states, scaled by the temperature of the system. If a physical system is energetically isolated then the work to change states is simply the change in the internal, or potential energies of those two states.

For arbitrary processes governed by transition rates L there may not be a natural definition of temperature, so the temperature may be defined as an arbitrary constant [5]. A natural constant to pick is $k_BT = 1$. Then the the edge flow is equal to one half the work required to cross each edge divided by a dummy variable with units of energy. In some situations it is natural to define a temperature based on system size, or based on the balance of diffusion to drift in the Markov chain. In those cases a different convention can be used to choose the scaling constant k_BT. In a physical model T should be the true temperature of the system.

A natural consequence of Equation (6.19) is that the sum of f over a path is one half the work to complete that path. Specifically, if $\{x_1, x_2, ...x_n\}$ is a sequence of states then the work required to traverse the sequence is [4]:

$$2\sum_{j=1}^{n-1} f_{x_{j+1}x_j} = \log\left(\prod_{j=1}^{n-1} \frac{l_{x_{j+1}x_j}}{l_{x_jx_{j+1}}}\right).$$

Notice that the product inside the logarithm is the ratio of the probability of traversing the

trajectory forward to the probability of traversing it backward, without the term associated with conditioning on the endpoints. If the path is a cycle then this is the ratio of the probability of traversing the cycle forward to traversing the cycle backward (see the discussion of reversibility in Section 6.3). Thus, with $k_B T = 1$, exponentiating the work to traverse a cycle recovers the ratio of the probability of traversing the cycle forward to traversing the cycle backward.

This gives a clean physical interpretation of the right hand side of the HHD, the edge flow represents the work to move between states, scaled by a dummy variable representing temperature. If we scale the potentials by the same dummy variables then they have units of energy. Then $-G\phi$ and $C^\intercal\theta$ both produce edge flows representing the work to move between edges. That is, the gradient of a potential function ϕ is work.

Usually the gradient of a potential is a force, since the gradient is usually a differential operator. In a discrete setting the gradient is a difference operator. If Δx_{ij} represents the change in state when moving from state i to state j, and $||\Delta x_{ij}||$ represents some measure of the magnitude of the change in state, then the average force over the transition from i to j would be $w_{ij}/||\Delta x_{ij}||$. Let D_x be a diagonal $E \times E$ weight matrix with diagonal entries equal to $1/||\Delta x_{i(k)j(k)}||$. Then $D_x w$ is the force required to cross each edge, and the rescaled gradient $D_x G$ is a finite difference approximation to a derivative operator since $[D_x G\phi]_k = \frac{\phi_{j(k)} - \phi_{i(k)}}{||\Delta x_{i(k)j(k)}||}$. Thus, if both sides of the HHD are be multiplied by D_x, then the rescaled gradient $D_x G$ of the potential recovers the conservative component of the forces. Note that this is a trivial reweighting of the HHD that does not change either ϕ or θ. In a purely abstract setting, with no state variables associated with the vertices, then moving from vertex i to vertex j corresponds to moving a unit probability mass, so it would be natural to assign each edge length one. Then the average force on each edge is equal to twice the edge flow, and the gradient of the potential is the force on each edge. Unless

there is a natural length associated with each edge (magnitude of the change in some state variable) we will use the convention that all edges have length one, so the edge flow equals the average force on each edge times a dummy variable with distance one.

If the system of interest obeys detailed balance then $f_{ji} = \phi_i - \phi_j$ for some scalar potential ϕ. If we rescale the potentials by $\frac{1}{k_B T}$ then $\phi_i - \phi_j = \frac{1}{k_B T} f_{ij} = \frac{1}{2} w_{ij}$. Therefore, if the system obeys detailed balance then the work to move from state i to state j can be expressed as the difference in a potential function, and the potential function at each vertex can be interpreted as the internal energy of the system when at those vertices. Alternatively, given an isolated physical system the work to move between states is always the difference in the internal energy of the states, so the edge flow obeys detailed balance. Then the scalar potential, after the appropriate scaling, would match the internal energy up to a factor of a half. If E_i is the interal energy of the i^{th} state then:

$$\phi_i = \frac{E_i}{2} \tag{6.20}$$

if the potentials are defined by $\frac{1}{k_B T}[-G\phi + C^\intercal \theta] = f$.

Equation (6.19) relates the HHD applied to the edge flow f to the energy and work associated with an analogous physical process. To build a complete thermodynamic analogy we also need to introduce entropy.

Entropy is a measure of the spread of a distribution, so is defined relative to a given probability distribution. Let $p(t)$ be the probability distribution for $X(t)$. Then, the entropy associated with the distribution p is [20, 185]:

$$H(p) = -\sum_{i \in \mathcal{V}} p_i \log(p_i) \tag{6.21}$$

where the base of the logarithm determines the units used to measure entropy. Changing

the base of the logarithm multiplies the entropy by a constant. In information theory the log is usually either base e (nats) or base 2 (bits) [185]. In a physical setting the natural units for entropy are given by using log base e, then multiplying by the Boltzmann constant k_B. Here entropy is denoted with H since this is the standard notation in information theory, and the definition of entropy provided by Equation (6.21) is the information theoretic definition of entropy [185]. It is important to note that the information theoretic definition of entropy and the statistical mechanical definition are not the same (differ by more than choice of units) for all systems [254]. The entropy is well defined for any probability distribution since, if $p_i = 0$, then $p_i \log(p_i)$ is set to 0 by convention since $\lim_{x \to 0} x \log(x) = 0$.

It is important to note that the entropy $H(p(t))$ does not necessarily increase over time [255]. A simple counterexample suffices.

Consider a network with two nodes. Then the network necessarily obeys detailed balance so the edge flow is the gradient of some potential. Assume that the potential at the two nodes are equal and the initial state, x_1 is known (i.e. $p(0) = [1, 0]$). Then, as time progress, $p_1(t)$ will decrease and $p_2(t)$ will increase until $p_1 = p_2$ and entropy is maximized. Suppose instead that the opposite is true. Let the potential at the first node be much much smaller than the potential at the second node. If the difference is taken towards infinity then the equilibrium distribution q approaches $[1, 0]$. Now start from $p(0) = [0.5, 0.5]$. The entropy of a distribution over N states is maximized by the uniform distribution $p = [1/N, 1/N ... 1/N]$ [185], so $H(p(0)) > H(p')$ for any $p' \neq p(0)$. Since $q \neq p(0)$ the entropy must decrease. In fact, since q is near to $[1, 0]$ the entropy is nearly minimized.

Clearly entropy $H(p)$ need not increase as time progresses. It is certainly possible that the equilibrium distribution of a Markov chain is more tightly distributed than the initial distribution in which case entropy will decrease. In fact entropy is only guaranteed to

increase as time progresses if the equilibrium q is the uniform distribution [185, 255]. The fact that the information entropy $H(p)$ may decrease does not, however, violate the second law of thermodynamics, since considering entropy alone ignores the energy associated with states, and the work to move between states. In order to consider entropy production properly we need to consider both the change in the entropy $H(p(t))$ and the energetics associated with the potentials and edge flow.

Suppose that the process of interest obeys detailed balance. Then the internal energy of the system can be expressed as the expected value of the potential:

$$\mathbb{E}[E_{X(t)}] = \sum_{i \in \mathcal{V}} E_i p_i(t) = 2 \sum_{i \in \mathcal{V}} \phi_i p_i(t). \tag{6.22}$$

Then the associated free energy $F(p)$ is defined [20]:

$$F(p(t)) = \mathbb{E}[E_{X(t)}] - TH(p(t)). \tag{6.23}$$

The free energy defined by Equation (6.23) is analogous to the Helmholtz free energy. However, since the number of nodes in the network is fixed the network volume is fixed. Similarly the total amount of probability in the network is conserved. Thus there is no meaningful analogy to pressure, so the Helmholtz and Gibbs free energies are equivalent up to the addition of a constant. Free energy may be interpreted as the energy still available in a system to do work. Isolated physical systems move to minimize their free energy. For example, the concentrations of chemical species in an isolated system of chemical reactions in an aqueous solution will eveolve to minimize their Gibbs free energy.

If we adopt the convention that $k_B T = 1$ then $TH(p(t)) = -k_B T \sum_{i \in \mathcal{V}} p_i(t) \log(p_i(t))$ which equals $-\sum_{i \in \mathcal{V}} p_i(t) \log(p_i(t))$ is the standard definition of entropy using nats. Then,

in terms of the potential ϕ:

$$F(p) = \sum_{i \in \mathcal{V}} (2\phi_i + \log(p_i))p_i. \tag{6.24}$$

Since we assumed that the process obeys detailed balance the scalar potential is directly related to the steady state distribution q by a Boltzmann type distribution (see Equation (6.10)). In particular $2\phi_i = -\log(q_i) - \log(Z)$ where Z is a normalization constant so that $\exp(-2\phi)/Z$ is normalized. Then:

$$F(p) = \left[\sum_{i \in \mathcal{V}} \log\left(\frac{p_i}{q_i}\right) p_i \right] - \log(Z) = D_{KL}(p||q) - \log(Z) \tag{6.25}$$

where $D_{KL}(p||q)$ is the Kullback-Liebler divergence (KL divergence) between the distribution p and the steady state distribution q [185, 254].

The KL divergence is the measure of the distance between two distributions. The KL divergence of p given q is the relative entropy of the distribution p relative to the distribution q [185]. In some closed physical systems the free energy can be shown to be equivalent to the relative entropy [254], so while the approach used here is entirely phenomenological, for appropriate physical systems (i.e. proteins and macromolecules in aqueous solution at constant temperature or polymer chains) the free energy defined by Equation (6.25) is the true free energy of the system.

The KL divergence is well defined for any p since, under the assumption of microscopic reversibility the Markov process is ergodic and has some chance of visiting any state, so the steady state distribution is not zero at any state. Note that the KL divergence is not a true distance since it is not symmetric in its argument and does not satisfy the triangle inequality [185].

The KL divergence is an example of an f-divergence:

$$D_f(p||p') = \sum_{i \in \mathcal{V}} p'_i f\left(\frac{p_i}{p'_i}\right) \tag{6.26}$$

where f is a convex function such that $f(1) = 0$. Any f-divergence defines a quasi-metric on the space of probability distributions since any f-divergence is nonnegative and zero if and only if $p = p'$, but generally is not symmetric in p and p', and generally does not satisfy the triangle inequality. The KL divergence is the f-divergence with $f(x) = x\log(x)$. Renyi introduced f-divergences in [256] where he showed that any f-divergence decreases for Markov processes. Since this fact is essential for the development of our thermodynamic analogy we review a proof provided by [185] for KL divergences.

The non-negativity of KL divergences is guaranteed by Jensen's inequality. If f is a convex function then Jensen's inequality states that $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ for any random variable X distributed according to any probability distribution. The f-divergence (Equation (6.26)) is $D_f(p||p') = \mathbb{E}[f(p(X)/p'(X))]$ where X is distributed according to p'. Therefore $D_f(p||p') \geq f(\mathbb{E}[p(X)/p'(X)])$. But:

$$\mathbb{E}[p(X)/p'(X)] = \sum_{i \in \mathcal{V}} \frac{p_i}{p'_i} p'_i = \sum_{i \in \mathcal{V}} p_i = 1.$$

By definition $f(1) = 0$ so $D_f(p||p') \geq f(\mathbb{E}[p(X)/p'(X)]) \geq f(1) = 0$. If $p = p'$ then $f(p'_i/p_i) = 0$ for all i so $D_f(p||p') = 0$. Thus any f-divergence is nonnegative and equals zero if $p = p'$.

The monotonicity of the KL divergence when applied to a Markov chain can be proved using the chain rule for relative entropy, the nonnegativity of f-divergences, and the Markov property.

Suppose X and Y are a pair of random variables with joint distributions $p(x, y)$ and $p'(x, y)$, marginal distributions $p(x)$ and $p'(x)$, and conditional distributions $p(y|x)$, $p'(y|x)$. Then the chain rule for relative entropy states that [185]:

$$D_{KL}(p(x, y)||p'(x, y)) = D_{KL}(p(x)||p'(x)) + D_{KL}(p(y|x)||p'(y|x)).$$

The chain rule can be proved by some trivial manipulation:

$$\begin{aligned}
D_{KL}(p(x, y)||p'(x, y)) &= \sum_{x,y} p(x, y) \log \left(\frac{p(x, y)}{p'(x, y)} \right) \\
&= \sum_{x,y} p(x, y) \log \left(\frac{p(x)p(y|x)}{p'(x)p'(y|x)} \right) \\
&= \sum_{x,y} p(x, y) \log \left(\frac{p(x)}{p'(x)} \right) + \sum_{x,y} p(x, y) \log \left(\frac{p(y|x)}{p'(y|x)} \right) \\
&= \sum_{x} p(x) \log \left(\frac{p(x)}{p'(x)} \right) + \sum_{x} p(x) \sum_{y} p(y|x) \log \left(\frac{p(y|x)}{p'(y|x)} \right) \\
&= D_{KL}(p(x)||p'(x)) + D_{KL}(p(y|x)||p'(y|x)).
\end{aligned}$$

Now suppose that $p(t), p'(t)$ are two different distributions representing the state of the Markov process initialized from possibly different initial distributions. Let $p(t+\Delta t), p'(t+\Delta t)$ be the distributions after time Δt has passed. Then $p(t + \Delta t) = \exp(L\Delta t)p(t)$ and $p'(t + \Delta t) = \exp(L\Delta t)p'(t)$ where $\exp(L\Delta t)$ is the discrete time transition matrix corresponding to the time interval Δt. This is a stochastic matrix, so its i^{th} column is the conditional distribution for the state of $X(t + \Delta t)$ given $X(t) = x_i$. Note that the same transition matrix updates both distributions.

Now let $p(x(t), x(t + \Delta t))$ denote the joint probability that $X(t) = x(t)$ and $X(t + \Delta t) = x(t + \Delta t)$. Marginal and conditional notation follows similarly. Then, using the

chain rule for relative entropy twice, once forward in time and once backward:

$$D_{KL}(p(x(t), x(t+\Delta t))||p'(x(t), x(t+\Delta t)))$$
$$= D_{KL}(p(x(t))||p'(x(t))) + D_{KL}(p(x(t+\Delta t)|x(t))||p'(x(t+\Delta t)|x(t)))$$

and

$$D_{KL}(p(x(t), x(t+\Delta t))||p'(x(t), x(t+\Delta t)))$$
$$= D_{KL}(p(x(t+\Delta t))||p'(x(t+\Delta t))) + D_{KL}(p(x(t)|x(t+\Delta t))||p'(x(t)|x(t+\Delta t)))$$

Now, since both $p(t+\Delta t)$ and $p'(t+\Delta t)$ are updated by the same transition matrix, the conditional distributions $p(x(t+\Delta t)|x(t))$ and $p'(x(t+\Delta t)|x(t))$ are the same, so their KL divergence is zero. Then, setting the two equations equal to one another, and moving $D_{KL}(p(x(t+\Delta t))||p'(x(t+\Delta t)))$ to the left hand side:

$$D_{KL}(p(x(t))||p'(x(t))) - D_{KL}(p(x(t+\Delta t))||p'(x(t+\Delta t)))$$
$$= D_{KL}(p(x(t)|x(t+\Delta t))||p'(x(t)|x(t+\Delta t))) \geq 0.$$

But $p(x(t)) = p(t)$ and $p'(x(t)) = p'(t)$ so:

$$D_{KL}(p(t)||p'(t)) \geq D_{KL}(p(t+\Delta t)||p'(t+\Delta t)) \text{ for any } \Delta t > 0. \tag{6.27}$$

Therefore the KL divergence between $p(t)$ and $p'(t)$, each governed by the master equation, but possibly initialized from different distributions, is monotonically decreasing.[4] That is, as time progresses distributions become more similar as they converge to the steady

[4]Note that this proof did not rely on detailed balance, so the KL divergence between $p(t)$ and its steady state q is always monotonically nonincreasing, so is a Lyapunov function. That said the interpretation of the KL divergence between $p(t)$ and q as the Helmholtz free energy depends on the relation between ϕ and q which requires detailed balance.

state. If $p'(0) = q$ then $p'(t) = q$ for all time since q is the steady state. Therefore the KL divergence between $p(t)$ and the steady state q is monotonically nonincreasing. It follows that the Helmholtz energy associated with a process that obeys detailed balance is monotonically nonincreasing in time:

$$\frac{d}{dt}F(p(t)) = \frac{d}{dt}\sum_i (2\phi_i + \log(p(t)_i))p(t)_i = D_{KL}(p(t)) - \log(Z) \leq 0. \tag{6.28}$$

Moreover, since $p(t)$ converges to q as t goes to infinity the KL divergence converges to zero, so the Helmholtz energy converges to $-\log(Z)$. Since the KL divergence is nonnegative this is the minimum possible value of the Helmholtz energy. Since the KL divergence is convex in p and p' [185], the free energy is a convex function in the distribution p.

Therefore, for a process obeying detailed balance, if the Helmholtz free energy is defined by Equation (6.23), or, equivalently, by Equation (6.24), then the free energy is a convex function in the distribution p, is nonincreasing in time, and $F(p(t))$ converges to its minimum value as $p(t)$ converges to the steady state. Thus the Helmholtz energy is a Lyapunov function[5] for the process $p(t)$ [5, 240]. The fact that all f-divergences have the same monotonic property means that any generalized free energy of the form $D_f(p||q)$ is a Lyapunov function for $p(t)$.

The monotonicity of the Helmholtz energy offers an elegant physical interpretation. Since the Helmholtz energy is the expected internal energy minus the entropy, the probability distribution moves simultaneously to minimize its energy and to maximize its entropy. If the system was purely deterministic the entropy would play no role, and the system would simply move to minimize its potential energy by aggregating in the node with the lowest energy. This is equivalent to excluding the diffusion term from the underlying stochastic

[5] A function $f(p)$ that is monotonically nonincreasing in $f(p(t))$ and reaches its minimum as t goes to infinity

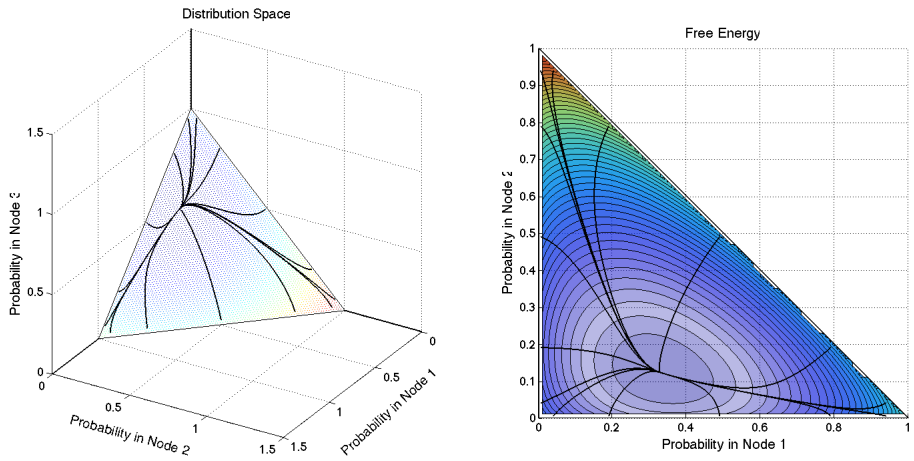

Figure 6.2: The left panel shows the space of possible probability distributions. The right panel shows the same subspace as viewed from above. The black curves represent paths taken by $p(t)$ as probability diffuses from initial conditions near the outer edges of the triangle. The colored contour plot on the right represent the Helmholtz Free Energy given $\phi = [0.5, 1, 0.25]$, with smaller values shown in purple and larger values in red. Notice that the Helmholtz free energy is a Lyapunov function for $p(t)$, and is minimized at equilibrium.

process. Since the system is not purely deterministic it gradually converts available energy into heat. This corresponds to the gradual loss of information. Entropy acts to spread the distribution uniformly, while energy acts to collapse the distribution towards some minimal states. The equilibrium distribution is a compromise between these two tendencies. Note that this mimics the usual properties of the free energy for systems of chemical reactions. In a system of chemical reactions the probability vector is replaced with concentrations, the free energy is decreasing as the concentrations change, and converges to a minimal value as the concentrations approach their steady states. The Helmholtz free energy for a three state network is shown in Figure 6.2, along with solution trajectories of $p(t)$.

The monotonicity of the Helmholtz energy is the second law of thermodynamics for Markov processes that obey detailed balance [20, 185]. Since energy is conserved, when-

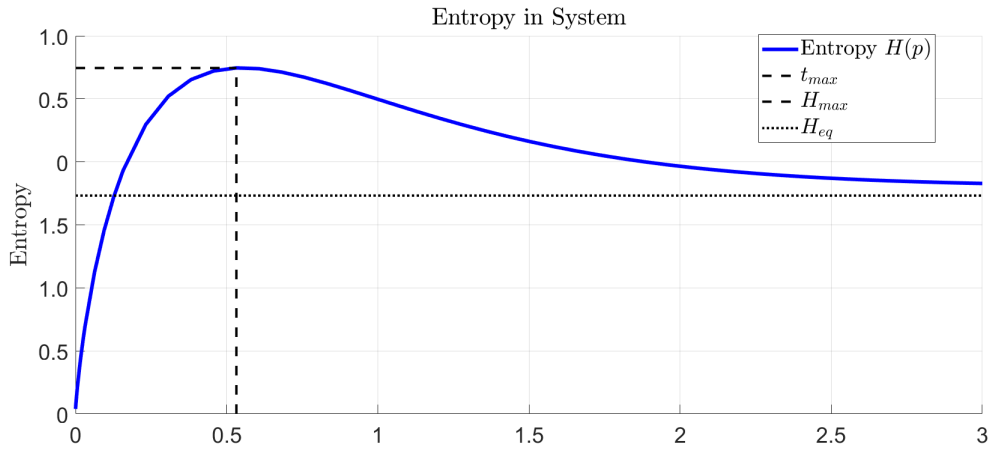

Figure 6.3: Entropy in a three state Markov chain increasing and decreasing with time.

ever the system moves between states with different energies that energy must be exchanged with the environment, typically by radiating heat. The radiated heat produces entropy in the environment, so the overall entropy production of the entire process is the change in internal entropy plus the entropy introduced into the environment by the release of heat [5]. Therefore, the rate at which the system produces entropy is the rate of decrease in the Helmholtz free energy [20]. Then the total entropy production is always nonnegative since the Helmholtz energy never increases. It follows that, if the internal entropy $H(p)$ decreases, then the decrease in entropy inside the system must be offset by a greater decrease in the internal energy, which releases enough heat into the environment.

Consider three nodes connected in a line. Suppose that node 1 is connected to node 2 is connected to node 3. Now suppose the flow along each edge point strongly away from the second node. Then the associated Laplacian could be:

$$L = \begin{bmatrix} -\epsilon & 1 & 0 \\ \epsilon & -2 & -\epsilon \\ 0 & 1 & \epsilon \end{bmatrix}$$

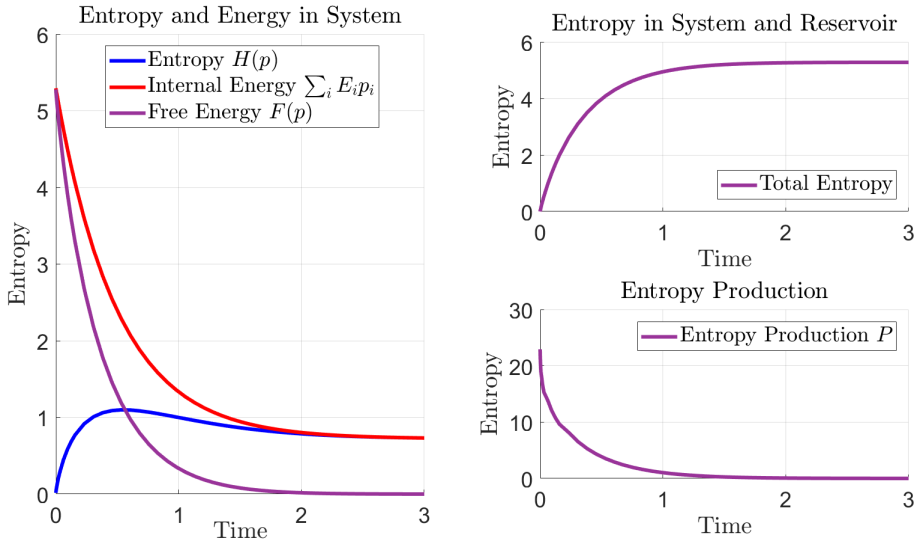

Figure 6.4: Entropy in a three state Markov chain (shown in blue), along with the expected internal energy (shown in red), and the free energy (shown in purple). The entropy increases and decreases, but the free energy decreases monotonically. The total entropy production ($-\frac{d}{dt}F(p(t))$) is shown in the bottom right. It is always positive, so the total entropy, shown in the top right, is always increasing.

where $\epsilon \ll 1$. Then the equilibrium q is approximately $[0.5, 0, 0.5]$. Set $p(0) = [0, 1, 0]$. Then the entropy at time zero is minimized.

When released, probability diffuses rapidly out of the second node, and accumulates in nodes 1 and 3. Since the process is symmetric there is some time t_{max} such that $p(t_{max}) = \frac{1}{3}[1, 1, 1]$ is uniform and the entropy is maximized $H(p(t_{max})) = H_{max} = -\log(1/3)$. At equilibrium q approaches $[0.5, 0, 0.5]$ which has entropy $H(q) \approx -\log(1/2)$. Since $H(p(0)) < H(q) < H(p(t_{max}))$ the entropy will first increase to a maximum, then decrease as the distribution approaches equilibrium. The corresponding process is shown in Figure 6.3.

While the entropy of the system, $H(p(t))$ decreases for times $t > t_{max}$ the free energy is monotonically decreasing, thus the total entropy produced by the system is increasing. This exchange of energy and entropy is shown in Figure 6.4

Following [5], let P denote the rate of entropy production. Then the entropy production is $-\frac{d}{dt}F(p(t)) = -\frac{d}{dt}\mathbb{E}[E_{X(t)}] + \frac{d}{dt}TH(p(t))$. In general, the time derivative of p can be expressed as the negative divergence of the probability fluxes, $\frac{d}{dt}p(t) = G^\intercal J(t)$. Therefore, if $f(p)$ is a scalar valued function of p, then by the chain rule, $\frac{d}{dt}f(p) = [\nabla f]^\intercal G^\intercal J(t) = (G[\nabla f])^\intercal J(t) = \sum_k (\partial_{p_{j(k)}} f(p) - \partial_{p_{i(k)}} f(p)) J_{i(k)j(k)}(t)$. That is, the rate of change of any function of the distribution is the same as the sum over the edges of the flux over each edge times the change in the function if an infinitesimal amount of probability is exchanged across the edge. Then:

$$\frac{d}{dt}TH(p(t)) = -\frac{d}{dt}\sum_i p_i(t)\log(p_i(t)) = \sum_{i<j}\left[(\partial_{p_i} - \partial_{p_j})\sum_h p_h(t)\log(p_h(t))\right] J_{ij}(t).$$

To simplify note that $\partial_{p_i}\sum_h p_h(t)\log(p_h(t)) = \log(p_i) + 1$ so:

$$\left[(\partial_{p_i} - \partial_{p_j})\sum_h p_h(t)\log(p_h(t))\right] = \log(p_i) - \log(p_j).$$

Therefore:

$$\frac{d}{dt}TH(p(t)) = \sum_{i<j}\log\left(\frac{p_i}{p_j}\right) J_{ij}(t). \tag{6.29}$$

It follows that $\log\left(\frac{p_i}{p_j}\right)$ is the infinitesimal change in the entropy associated with moving an infinitesimally small amount of probability between nodes i and j. If $p_j < p_i$ then the ratio is greater than one, so the entropy increases as probability flows from i to j. If $p_j > p_i$ then the ratio is less than one, so the entropy decreases as probability flows from i to j.

The rate of change in the internal energy is:

$$\frac{d}{dt}\mathbb{E}[E_{X(t)}] = \sum_{i<j} 2(\phi_i - \phi_j) J_{ij}(t). \tag{6.30}$$

But $-G\phi = f$ so $2(\phi_i - \phi_j) = 2f_{ij} = \log(l_{ij}) - \log(l_{ji})$. Therefore the total rate of entropy production is:

$$P = -\frac{d}{dt}\mathbb{E}[E_{X(t)}] + \frac{d}{dt}TH(p(t)) = \sum_{i<j} \log\left(\frac{l_{ij}p_j}{l_{ji}p_i}\right) J_{ij}(t). \tag{6.31}$$

The log ratio appearing before the probability flux J are the affinities [5].

Affinities generalize the notion of forces from classical mechanics to thermodyanmics. In classical mechanics forces are derivatives of energy functions, so are the change in energy after an infinitesimal motion. The affinity associated with an edge is the change in free energy associated with moving an infinitesimal amount of probability across the edge. Thus, in detailed balance the the affinities A_{ij} are given by:

$$A_{ij}(p) = -[\nabla F(p)](e_i - e_j) = 2(\phi_i - \phi_j) + \log\left(\frac{p_i}{p_j}\right) = \log\left(\frac{l_{ji}p_i}{l_{ij}p_j}\right). \tag{6.32}$$

That is, the affinities are the projection of the negative gradient of the Helmholtz energy onto the directions $e_i - e_j$ associated with the flow of probability along the edges. The first term in the affinity is the force on the edge ij associated with the work to move across the edge. The second term is a diffusive force associated with the tendency to maximize entropy.

Note that if the network consisted of a single edge then it would automatically obey detailed balance so $l_{ij}q_j$ would equal $l_{ji}q_i$, in which case $A(q) = 0$. More generally, if a process obeys detailed balance, then the affinity goes to zero at steady state. Mathematically this is a direct result of the detailed balance condition, since $l_{ji}p_i = l_{ji}p_j$ gives $\log\left((l_{ji}p_i)/(l_{ij}p_j)\right) = \log(1) = 0$. Thus, in detailed balance, steady state is achieved when all of the "force" at play in the system balance.

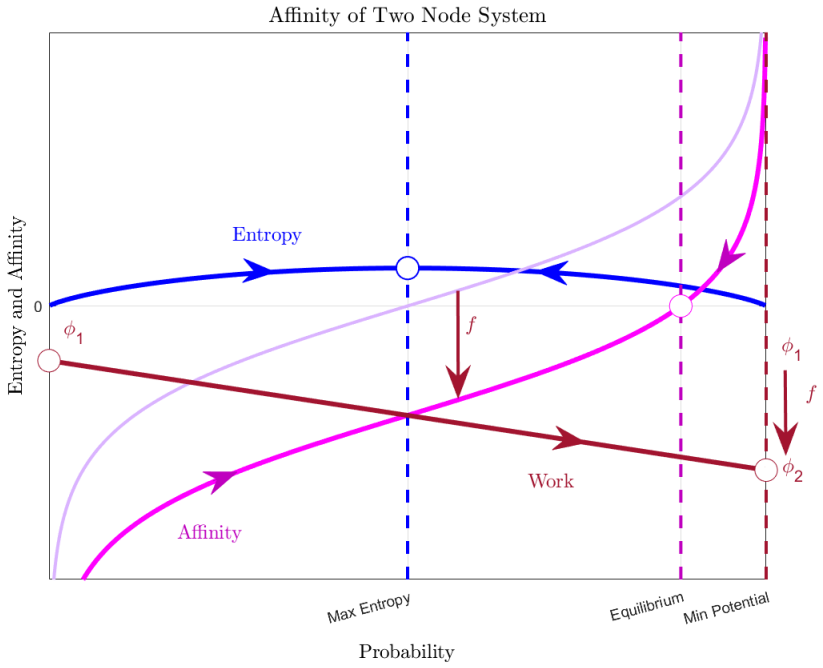

Figure 6.5: Affinity of a two state system with the associated potentials ϕ, edge flow f, and entropy H. The affinity A is shown in magenta, the potentials and associated forces in red, and the entropy in blue. The affinity given no edge flow is shown in light magenta, and is the partial derivative of the entropy corresponding to an infinitesimal exchange of probability between the nodes. The difference in potential over the edge introduces a force f which shifts the affinity down. The equilibrium distribution is the intersect of the affinity with zero. This equilibrium falls between the uniform distribution, which maximizes entropy, and the distribution $[0, 1]$ which minimizes the potential.

The affinity associated with a single edge is shown in Figure 6.5. The component of the affinity associated with the work to cross the edge, and the component associated with the entropy are shown separately. If the network only consisted of one edge then the steady state distribution is q such that $A_{ij}(q) = 0$. The corresponding equilibrium is shown in Figure 6.5.

Since the affinity is the change in free energy associated with an infinitesimal exchange of probability across each edge, the entropy production is given by the inner product [20,

253]:

$$P(p) = A(p)^\intercal J(p) = \sum_{i<j} A_{ij}(p) J_{ij}(p). \tag{6.33}$$

The entropy production, P, defined by Equation (6.33) is closely related to the Joule heating in the electric circuit analogy developed in Section 6.3.1. Recall that, if a network obeys detailed balance, then the transition rates l_{ij} and l_{ji} can be rewritten $\frac{1}{R_{ij}C_j}$ and $\frac{1}{R_{ji}C_i}$, where $R_{ij} = R_{ji}$ is the resistance on edge ij, and $C_j = q_j$ are capacitances and equal the steady state distribution. Then the affinities are:

$$A_{ij}(p) = \log\left(\frac{C_j}{p_j}\frac{p_i}{C_i}\right) = \log\left(\frac{p_i}{C_i}\right) - \log\left(\frac{p_j}{C_j}\right).$$

If the system is near steady state then $p_j \simeq C_j$, so each ratio is near one, and the logarithms can be replaced with their Taylor expansions:

$$A_{ij}(p) \simeq \frac{p_i}{C_i} - \frac{p_j}{C_j} + \mathcal{O}\left(\left(1 - \frac{C_j}{p_j}\frac{p_i}{C_i}\right)^2\right).$$

If the probabilities represent a charge distribution then the linear term is the voltage over the edge, so $A_{ij}(p) \simeq V_{ij}$ to first order.

The currents in the electric circuit analogy obey Ohm's law, so the entropy production near steady state is:

$$P(p) = \sum_{i<j} A_{ij}(p) J_{ij}(p) \simeq \sum_{i<j} V_{ij} J_{ij}(p) = \sum_{i<j} \frac{V_{ij}^2}{R_{ij}}.$$

The sum of the voltage drop squared over the resistance on every edge is the heat produced by the current over the edge [34]. Thus, to first order the entropy production of a process obeying detailed balance matches the Joule heating of the analogous electric

circuit. The fact that this approximation only applies near steady state is not surprising since Ohm's law and Joule heating are themselves linearizations that only approximate the true current and heat loss when the currents are relatively small.

Most of our discussion so far has focused on the detailed balance case, in which the physical system is energetically isolated. Not all Markov chains obey detailed balance, after all, if all Markov chains obeyed detailed balance then there would be no need for the HHD. If the Markov chain of interest does not obey detailed balance then the analogous physical system is coupled to an external energy source that drives circulation [5].

The relationship between the edge flow and work defined by Equation (6.19), and information theoretic definition of entropy still apply out of detailed balance. The affinities can also be defined in the same way, only with an additional term to account for work done by external energy sources. If the affinity on an edge is defined by the free energy used to move an infinitesimal amount of probability across the edge, then the affinity equals the work to cross the edge plus the change in entropy associated with moving an infinitesimal amount of probability across the edge. The work captures the total energy exchanged with the reservoir, so the work to cross the edge still accounts for entropy produced in the reservoir. The key difference when working with a system that does not obey detailed balance is that the work to cross an edge may not be given exclusively by the difference in potential on either side.

Let $w_{ij} = 2f_{ij}$ be the work to cross edge ij from i to j. Then the affinity on edge ij is [5]:

$$A_{ij}(p) = -w_{ij} + [\nabla H(p)](e_i - e_j) = \log\left(\frac{l_{ji}p_i}{l_{ij}p_j}\right) \tag{6.34}$$

where the second equality follows from the definition of the edge flow. Notice that the affinity defined by Equation (6.34) is the same as the affinity defined by Equation (6.32) if

the form using the log ratio of the transition rates is used. This is the form provided by [5].

We can use the network HHD to write the affinities in terms of the potentials directly:

$$A(p)_{ij} = -2f_{\text{con}_{ij}} - 2f_{\text{rot}_{ij}} + f_{\text{dif}_{ij}}(p) = -2[G\phi]_{ij} + 2[C^{\intercal}\theta]_{ij} + \log\left(\frac{p_i}{p_j}\right). \tag{6.35}$$

where f_{dif} stands for the diffusive forces associated with the entropy. Note that if the Helmholtz free energy is still defined, $F(p) = \sum_i p_i\phi_i + p_i \log(p_i)$ then the affinity can be rewritten:

$$A(p)_{ij} = 2(f_{\text{con}_{ij}} + f_{\text{rot}_{ij}}) + \log\left(\frac{p_i}{p_j}\right) = 2f_{\text{rot}_{ij}} - (\nabla F(p))^{\intercal}(e_i - e_j). \tag{6.36}$$

In this context the affinity on a given edge is a combination of a rotational force, and the partial derivative of the Helmholtz Free Energy along the edge.

Our main interest now is to associate the vector potential θ with external energy sources exerting work on the system. That is, to identify rotation driven by external forces that act on the system with the rotational potential.

To compute the net external forces driving rotation on a given loop, sum the affinities about the loop [5]. Any path integral around the loops of the network can be decomposed into path integrals about basis loops so we will consider the force necessary to drive rotation around each basis loop. Then the collection of net external forces is simply:

$$A_{ext} = CA(p) \tag{6.37}$$

where A_{ext} are the external forces, and $A(p)$ is the vector of affinities on each edge. Notice that the external forces are independent of the probability distribution inside the system [5]. This is an essential consistency condition. The forces are necessarily external, since, if

the process completes a cycle then any work done over the cycle must be associated with energy exchanged with the reservoir [4].

Expand the affinities to prove that the external forces are independent of the distribution p inside the system:

$$A_{ext} = CA(p) = C\left(-2G\phi + 2C^{*}\theta + f_{\text{dif}}(p)\right).$$

The curl is orthogonal to the gradient, so A_{ext} does not depend on ϕ. The diffusive force is the log of the ratio of probabilities, which is a geometric difference of a function p defined on the states. The curl of any function of this type is also zero. As a simple example consider a loop $i \to j \to k \to i$. The curl of the diffusion term is:

$$\log\left(\frac{p_i p_j p_k}{p_i p_k p_j}\right) = \log(1) = 0.$$

Therefore the set of external forces equals to the curl of the rotational field, or, the face Laplacian of the vector potential:

$$A_{ext} = 2Cf_{\text{rot}} = 2L_{\mathcal{C}}^{2}\theta. \tag{6.38}$$

Thus the affinity on any edge is a combination of a diffusion term, a force associated with the scalar potential that represents internal energy, and a force associated with the vector potential that represents coupling to energy sources in a reservoir. The first potential, ϕ, is related to the free energy of the system. The second potential is external, θ, and is related to the forces applied to the system that drive rotation about closed loops [4]. If we consider the component associated with the rotational potential as external then the generalized force is a combination of internal and external forces. The internal forces

arise directly from the gradient of the Helmholtz free energy, while the external forces are associated with the rotational component of the edge flow/work, f_{rot}.

The external affinities, A_{ext}, equal the work to traverse each basis cycle. Since the work to complete a cycle equals the ratio of the probability of completing the path forward to the probability of completing the path backward, the external affinities are immediately related to the long time limit of the number of times each cycle is traversed forward relative to backward [4]. For a cycle $\mathcal{C}$ let $N_{\mathcal{C}+}(t)$ be the number of times the process $X(t)$ has completed a forward traversal of the cycle $\mathcal{C}$ at time t. Similarly let $N_{\mathcal{C}-}$ be the number of completed backward traversals. Then let $J_{\mathcal{C}+}(t) = N_{\mathcal{C}+}(t)/t$ and let $J_{\mathcal{C}-}(t) = N_{\mathcal{C}-}(t)/t$. In order to complete a cycle the process must first arrive at a node in the cycle. Once $X(t)$ has arrived at a node in the cycle the ratio of the probability of completing the cycle forward instead of backward, conditioned on completing a cycle is exponential in the work to complete the cycle. Thus, if $J_{\mathcal{C}+} = \lim_{t\to\infty} J_{\mathcal{C}+}(t)$, and $J_{\mathcal{C}-} = \lim_{t\to\infty} J_{\mathcal{C}-}(t)$, with convergence in probability, then in the long time limit [4]:

$$\frac{J_{\mathcal{C}+}}{J_{\mathcal{C}-}} = \exp(2C(\mathcal{C})f) = \exp(A_{ext}(\mathcal{C})). \tag{6.39}$$

Therefore the external affinity on a loop is associated with the ratio of the number of forward traversals to backward traversals in the long time limit. Note that this is not enough to recover the steady state probability flux around the loop, CJ, since the flux around the loop is $J_+ - J_- = \sqrt{J_+J_-}(\sqrt{J_+/J_-} - \sqrt{J_-/J_+}) = \sqrt{J_+J_-}2\sinh(A_{\text{ext}}/2)$ and $\sqrt{J_+J_-}$ is not fixed by the work to complete the loop. The long time distribution of forward and backward cycle traversals is studied by [6].

If the affinities are defined by Equation (6.34) then the entropy production is still defined by the inner product of the affinities with the flux, since the affinity on an edge is the total

free energy used to move probability over the edge. Thus $P(p) = \sum_{i<j} A_{ij}(p) J_{ij}(p)$ recovers the entropy production for systems both in and out of detailed balance [5].

If the process does not obey detailed balance then the steady state fluxes are nonzero. The steady state fluxes are necessarily divergence-free, so must be in the range of the curl. It follows that the steady state fluxes can be written $C^\intercal \theta_J$ for some $\theta_J \in \mathbb{R}^L$. Therefore, the steady state entropy production P depends only on the external forces applied to the system and the steady state flux:

$$P(q) = J^\intercal(q) A(q) = \theta_J^\intercal C A(q) = \theta_J^\intercal A_{ext} = 2\theta_J^\intercal C f_{\text{rot}} = 2 J^\intercal(q) f_{\text{rot}}. \tag{6.40}$$

That is, when a process does not obey detailed balance, the steady state entropy production is the inner product of the steady state fluxes with the rotational component of the HHD. Note that this can be reduced to an inner product defined on the space of loops rather than edges. Summing over a set of basis loops rather than the edges is often preferable since the long term production of observables is associated with the amount of observable produced per cycle completed, and the coupling to the reservoir is expressed through the tendency of external energy sources to drive circulation. Also note that we only need to sum over a set of L cycles rather than all possible cycles. This is an advantage of Equation (6.40) over the entropy production formula given by [4].

6.6 Summary

In this chapter we introduced discrete-space continuous-time Markov chains and showed that, for an appropriately chosen edge flow, the HHD is closely related to the thermodynamics of an analogous physical process. The detailed balance case (conservative edge flow) was highlighted (see Section 6.3). It was shown that in detailed balance the scalar

potential determines the steady state distribution. Outside of detailed balance the steady state is maintained by circulation of probability. An electric circuit analogy was introduced to build intuition in the detailed balance case. In Section 6.5 a general thermodynamic analogy was introduced which relates the time evolution of the probability distribution to a free energy function, work exerted to cross edges (the edge flow), and the internal entropy of the system. A thermodynamic interpretation of the edge flow (work) and potentials (energy) was introduced.

Chapter 7

Dynamics

7.1 Preface

In this chapter we analyze the dynamics of continuous-time discrete-space Markov processes using the HHD. We focus on nonequilibrium steady states and steady state fluxes, and give exact limiting expressions for nonequilibrium steady states in terms of the HHD. We show that the long term production rate of any observable based on path integration can be simplified using the HHD, and that the space of observables that are martingales in and near detailed balance are controlled by the ranges of the operators used in the HHD.

First we show that any nonequilibrium process can be transformed into a purely rotational process by scaling the Laplacian by the steady state distribution associated with the corresponding equilibrium process (see Section 7.2.1). This transform reduces the problem of solving for an arbitrary nonequilibrium steady state to solving for the steady state of purely rotational processes. Then, in Section 7.3.1 and Section 7.3.2 we show that, if the Markov chain is dominated by diffusion (small edge flow, weak forcing), then

nonequilibrium steady states and steady state fluxes can be approximated using the HHD. A formal expansion of the steady state and steady state fluxes in the strength of rotation is introduced in Section 7.3.2, and it is shown that, at every order, the terms in the expansion are solutions to a recursive sequence of weighted HHD equations. We show that the first order terms in this expansion are closely related to linear thermodynamics (thermodynamics for processes near equilibrium). We also explore what requirements on the network topology and conductances are needed to ensure that the steady state is rotation independent (see Section 7.3.2).

In the opposite limit, when drift dominates diffusion (large edge flow, strong forcing), then a different potential framework is needed (see Section 7.3.3). This framework is analogous to the quasipotential used to study SDE's in a small noise limit.

7.2 Nonequilibrium Steady States

7.2.1 The Purely Rotational Transform

To begin, suppose that L obeys detailed balance. Then there exists a set of resistances and capacitances so that the probability flux can be written:

$$J_{ij}(p) = \frac{1}{R_{ij}}\left(\frac{p_i}{C_i} - \frac{p_j}{C_j}\right) \tag{7.1}$$

where the resistances $R_{ij} = R_{ji}$ are one over the rate at which probability is exchanged between i and j at equilibrium ($\rho_{ij}\sqrt{q_i q_j} = \sqrt{l_{ji}q_i l_{ij}q_j} = l_{ji}q_i = l_{ij}q_j$), and the capacitances C_i equal the steady state occupancy of each node, q_i (see Section 6.3.1). If the scalar potential is shifted appropriately then $C_i = q_i = \exp(-2\phi_i)$.

The rate of change of probability at any node is the rate at which probability flows into the node minus the rate it flows out, which is the (negative) divergence of the probability flux. The negative divergence is just the gradient transpose. The rate of change of probability is also governed by the Laplacian through the master equation (see Equation (6.2)). Therefore, when in detailed balance, the Laplacian can be decomposed:

$$L = G^{\intercal}\left[-R^{-1}GQ^{-1}\right] = -\left[G^{\intercal}R^{-1}G\right]Q^{-1} \tag{7.2}$$

where R is a diagonal $E \times E$ matrix with diagonal entries set to the resistances on the edges and Q is a diagonal $V \times V$ matrix with diagonal entries equal to $C = q = \exp(-2\phi)$. This can be seen directly by noting that $l_{ij} = \rho_{ij}\exp\left(\phi_j - \phi_i\right) = \rho_{ij}\exp\left(-[\phi_j + \phi_i]\right)\exp\left(-2\phi_j\right) = R_{ij}^{-1}q_j^{-1}$ and $l_{jj} = -\sum_{i\in\mathcal{N}(j)}\rho_{ij}\exp\left(\phi_j - \phi_i\right) = -\sum_{i\in\mathcal{N}(j)}R_{ij}^{-1}q_j$.

Equation (7.2) can be read in two different ways. First, the Laplacian is the negative divergence of a matrix which maps from the probabilities p to the fluxes $J(p) = [R^{-1}GQ^{-1}]p$. Second, the Laplacian is the product of a symmetric matrix $G^{\intercal}R^{-1}G$ with a diagonal matrix Q^{-1} that scales the probability at each node relative to the steady state probability. In the electric circuit interpretation the product $Q^{-1}p$ produces the voltage at each node relative to the ground. Then the matrix $G^{\intercal}R^{-1}G$ is the matrix responsible for mapping from voltages to change in charge.

Let:

$$\begin{aligned}\hat{p} &= Q^{-1}p \\ \hat{L} &= LQ = -G^{\intercal}R^{-1}G.\end{aligned} \tag{7.3}$$

Then $\frac{d}{dt}p = \hat{L}\hat{p}$. Since the probabilities are all nonnegative and the steady state is strictly positive at every node $\hat{p}$ is also nonnegative. Therefore, if scaled by an appropriate normalizing constant $\hat{p}$ can be interpreted as a probability distribution. Then $p_i = \frac{1}{Z}q_i\hat{p}_i$

where Z is the necessary normalization. This expresses the probability distribution for the original unscaled process (Equation (7.1)) relative to the steady state.

The main advantage to this approach is that $\hat{L}$ is still interpretable as the Laplacian for a continuous-time discrete-space Markov chain on the same network, but $\hat{L}$ is symmetric, so corresponds to a weighted random walk. This relationship is easy to prove.

Consider a weighted random walk on $\mathcal{G}_{\rightleftarrows}$ with symmetric forward and backward transition rates $w_k = w_{i(k)j(k)} = w_{j(k)i(k)}$. Then the corresponding Laplacian can be constructed from the weights by performing the product $G^\intercal W G$ where $W_{kk} = w_k$. This can be seen constructively. The ij entry of the product $[G^\intercal W G]_{ij}$ is the weighted inner product between the i^{th} column of G and the j^{th} column of G. These columns correspond to node i and j, and the rows of the columns correspond to the edges. The rows of column i are only nonzero at edges that connect to node i, and the rows of column j are only nonzero at edges which connect to node j. The nonzero entries are all equal to ± 1. Therefore, if $i = j$ the weighted inner product is the sum of the weights on all edges neighboring node i. If $i \neq j$ but i and j are connected then the only row where both columns are nonzero is the row corresponding to the edge ij. One entry must be $+1$ and the other -1 so $[G^\intercal W G]_{ij} = -w_{ij}$ if nodes i and j are connected. If they are not connected then the two columns share no nonzero entries in common so the product is zero. Then:

$$-[G^\intercal W G]_{ij} = \left\{ \begin{array}{l} w_{ij} \text{ if } i \neq j \text{ and there is an edge between them} \\ 0 \text{ if } i \neq j \text{ and there is not an edge between them} \\ -\sum_{k \in \mathcal{N}_i} w_{ik} \text{ if } i = j \end{array} \right\} \tag{7.4}$$

which is exactly the structure of the Laplacian for a simple random walk with weights W.

Therefore $\hat{L} = -G^\intercal R^{-1} G$ is the Laplacian for the simple random walk with weights

set to one over the resistances. It follows that if $p(t)$ is known at some time, then $\hat{p}(t)$ can be computed by scaling by the steady state and normalizing. Then $\hat{p}$ obeys the master equation corresponding to a simple random walk with weights set to one over the resistances. Since master equation conserves probability $\sum_i \hat{p}(t) = 1$ for all t, so after the initial rescaling $\hat{p}(t)$ behaves exactly like the probability distribution for a simple random walk with weights equal to one over the resistances.

Thus, after rescaling by the steady state distribution and a normalization constant, the original detailed balance process can be transformed into a simple random walk with a symmetric Laplacian whose transition rates are one over the resistances of the original network.

This result can be generalized to arbitrary Laplacians in order to transform a generic nonequilibrium Markov process into a purely rotational process.

Lemma 25 (The purely rotational transform)**.** *Let $X(t)$ be a continuous-time discrete-space Markov process that obeys microscopic reversibility on a connected graph $\mathcal{G}_{\rightleftarrows}$, and with transition rates L. Let ρ denote the per capita conductances be $\rho_{ij} = \sqrt{l_{ij} l_{ji}}$. Let f denote the edge flow $f_{ij} = \frac{1}{2}\left(\log(l_{ji}) - \log(l_{ij})\right)$.*

Let ϕ denote the scalar potential associated with the HHD of f, and $\mathrm{diag}(\exp(-2\phi))$ *denote the diagonal matrix with entries equal to* $\exp(-2\phi_i)$*. Then:*

$$\hat{L} = L\mathrm{diag}(\exp(-2\phi)) \tag{7.5}$$

is the Laplacian for a Markov process whose edge flow is purely rotational with conductances equal to the rate of probability transfer between nodes at the equilibrium of the

original process with the rotational component of the flow removed. Moreover:

$$\begin{aligned} p_i(t) &= \frac{1}{Z}\exp(-2\phi_i)\hat{p}_i(t) \\ \frac{d}{dt}p(t) &= \frac{1}{Z}\hat{L}\hat{p}(t) \end{aligned} \tag{7.6}$$

where the normalization constant Z is chosen so that $\sum_i \hat{p}_i(t) = 1$, and ϕ is shifted so that $\sum_i \exp(-2\phi_i) = 1$. Therefore if q is the steady state of the original process and $\hat{q}$ is the steady state of the scaled process, then $q = \frac{1}{Z}\exp(-2\phi_i)\hat{q}_i$.

Proof. The product $L\text{diag}(\exp(-2\phi))$ scales the rows of L. Expand L into the conductances and edge flow according to Equation (6.16). Then the product for a particular off-diagonal entry is:

$$\hat{l}_{ji} = l_{ji}\exp(-2\phi_i) = \rho_{ij}\exp(f_{ij} - 2\phi_i).$$

Note that the indexing convention is reversed for the Laplacian so that the transition rate from i to j is indexed ji. This indexing ensures that the Laplacian can be used in the master equation without a transpose. Then, expanding f using the HHD:

$$f_{ij} - 2\phi_i = f_{\text{con}_{ij}} + f_{\text{rot}_{ij}} - 2\phi_i = \phi_i - \phi_j + f_{\text{rot}_{ij}} - 2\phi_i = -(\phi_i + \phi_j) + f_{\text{rot}_{ij}}.$$

Therefore:

$$\hat{l}_{ji} = [\rho_{ij}\exp(-\phi_i - \phi_j)]\exp(f_{\text{rot}_{ij}}).$$

The term in brackets is symmetric in i and j, so can be interpreted as a weight, or one over a resistance. In fact, if the rotational component of the flow was removed from the original process, then it would have obeyed detailed balance and $[\rho_{ij}\exp(-\phi_i - \phi_j)]$ would equal one over the resistance on edge ij, $1/R_{ij}$. In turn, $1/R_{ij}$ equals the rate at which

probability is exchanged between nodes i and j at the equilibrium (see Section 6.3.1). Each entry of L is scaled by a positive number to produce $\hat{L}$ so all of the off-diagonal entries of $\hat{L}$ are nonnegative.

The diagonal entries of the product are $\hat{l}_{ii} = -[\sum_{j \in \mathcal{N}_i} l_{ji}] \exp(-2\phi_i) = -\sum_{j \in \mathcal{N}_i} \hat{l}_{ji}$ so the columns of $\hat{L}$ all sum to zero. It follows that $\hat{L}$ is the Laplacian for a Markov process on $\mathcal{G}_{\rightleftarrows}$ with conductances set to the rate at which probability moves between nodes at the equilibrium of the original process without its rotational component, and with edge flow equal to the rotational component of the original edge flow. It follows immediately that if the original process obeyed detailed balance then the transformed Laplacian $\hat{L}$ would correspond to a simple random walk with weights equal to one over the resistances.

Equation (7.6) follows directly from the observation that $\exp(2\phi_i)$ are all positive, and $p(t)$ are all nonnegative, so all the entries of $\hat{p}(t)$ are nonnegative, and for the appropriate choice of Z the distribution $\hat{p}$ is normalized. Then $\frac{1}{Z}\hat{l}_{ji}\hat{p}_i = l_{ji}\frac{1}{Z}\exp(-2\phi_i)\hat{p}_i = l_{ji}p_i(t)$.

If $\hat{q}$ is the steady state for the scaled process then $\hat{L}\hat{q} = 0$ so $\frac{d}{dt}p(t) = 0$ if $p = \frac{1}{Z}\exp(-2\phi)\hat{q}$. □

Thus, if the Laplacian of an arbitrary nonequilibrium process is rescaled by the equilibrium distribution of the corresponding equilibrium process, $q^{(eq)} = \frac{1}{Z}\exp(-2\phi)$, then the resulting Laplacian is purely rotational, with edge flow equal to the rotational component of the original edge flow, and conductances equal to the rate at which probability is exchanged between nodes at the equilibrium distribution corresponding to the conservative component of the original process. This transform is powerful since it reduces the general problem of finding a nonequilibrium steady state, with an arbitrary combination of conservative and rotational components, to the special case when the Markov process is purely rotational.

Note that the steady state fluxes of the purely rotational problem (up to a rescaling by a normalization constant) are the steady state fluxes of the original problem since the fluxes

satisfy:

$$\hat{J}_{ij} = \hat{l}_{ij}\hat{q}_j - \hat{l}_{ji}\hat{q}_i = \hat{l}_{ij}\frac{{q^{(eq)}}_j}{{q^{(eq)}}_j}\hat{q}_j - \hat{l}_{ji}\frac{{q^{(eq)}}_i}{{q^{(eq)}}_i}\hat{q}_i = l_{ij}Zq_j - l_{ji}Zq_i = ZJ_{ij}$$

so the steady state fluxes after the transform are the steady state fluxes of the original system.

7.2.2 Isolated Cycles

The simplest rotational system is a network with isolated (edge disjoint) cycles. That is a network whose cycles do not share any edges. The cycles may be singly connected, so pairs of cycles may share a node, or may be connected by paths that include edges that are not part of any cycle. Since the cycles are isolated the steady state on each can be studied in isolation.

In this section we solve for the steady state for an arbitrary network with isolated loops. The result provides helpful intuition for the weak and strong rotation limits. Suppose $\mathcal{G}$ has a set of loops that are all edge disjoint. Assume that the network has potentials ϕ and θ and conductances ρ.

First, rescale by $\exp(-2\phi)$ to get the purely rotational system $\hat{L}$ from L. This transforms the problem into a purely rotational problem with a new set of conductances $\hat{\rho}_{ij} = \rho_{ij}\exp(-(\phi_i + \phi_j))$. The true steady state can be recovered by scaling the steady state of the purely rotational problem by the equilibrium distribution $\exp(-2\phi)$.

Since it is always possible to scale to a purley rotational system we start by considering the steady state of the purely rotational system. Unless the original system obeyed detailed balance the new steady state will be dynamic - there will be nonzero fluxes on some edges. Then the steady state is maintained by circular balance [4], where the net flux into and out

of each node is zero. Since we assumed that each loop is disjoint, there cannot be any flux across edges connecting loops that are not part of a loop themselves. Suppose that there was an edge, not included in any cycle, with a nonzero steady state flux across it. Then that edge would be a cut edge for the graph, so removing the edge from the graph would break it into two components. If there is a nonzero flux across the edge then the total probability in one component must be increasing, and the total probability in the other must be decreasing. The probabilities must be constant at steady state so it is impossible for there to be any steady state flux across an edge not included in a cycle. Moreover, if edges ij and jk are in the same loop then $J_{ij} = J_{jk}$ since the divergence of the fluxes must be zero at steady state. Therefore the flux across all edges in a given loop is constant.

Our goal now is to work out the steady state distribution and steady state fluxes on an individual loop. The full steady state can be constructed by piecing together steady state distributions for each individual loop. This is accomplished by noting that, if an edge is not in a loop then the steady state flux across the edge must be zero, and there is no rotational force on the edge. If there is no rotational force on the edge then there are no forces on the edge, so the forward rate and backward rate are identical. Therefore the steady state distribution is identical at either end of the edge. So, if two loops are connected by a path, then the steady state distribution over the path is constant, and the nodes at the ends of the path in the two loops must have the same steady state probability. Given two individual steady states, one for each loop independently, we can always shift and rescale so that both have the same probability at a pair of nodes. Therefore the full steady state can be built by matching the independent steady states of each loop at nodes which are connected by paths not included in any loop.

This reduces the problem of solving for a general steady state to solving for the steady state on single loops. Focus on a particular loop. Denote the steady state flux on that loop

J. Index all the nodes from 1 to $|\mathcal{C}|$ where $|\mathcal{C}|$ is the perimeter of the loop, and the node indices increase in the direction of rotation. Index the edges so that edge k points in the direction of rotation from k to $k+1$. Since all the loops are disjoint $f_{\text{rot}_k} = \theta$ on every edge in the loop. Moreover, since the loop is isolated θ is just the curl of f on the loop, divided by its perimeter $|\mathcal{C}|$. Therefore the forward and backward rates are:

$$l_k^+ = \frac{1}{R_k}\exp(\theta), l_k^- = \frac{1}{R_k}\exp(-\theta). \tag{7.7}$$

where R_k, the resistance of k^{th} edge, and θ are defined by:

$$R_k = \hat{\rho}_k^{-1} = \frac{\exp(\phi_k + \phi_{k+1})}{\sqrt{l_k^+ l_k^-}}, \quad \theta = \frac{\sum_{k=1}^{|\mathcal{C}|} \log(l_k^+) - \log(l_k^-)}{2|\mathcal{C}|}. \tag{7.8}$$

Let $\alpha = \exp(\theta)$. Then the edge rates are simply:

$$l_k^+ = \frac{\alpha}{R_k}, l_k^- = \frac{1}{\alpha R_k}.$$

Then the steady state flux on the edge is:

$$\frac{\alpha}{R_k} q_k - \frac{1}{\alpha R_k} q_{k+1} = J$$

where the nodes are counted modulo $|\mathcal{C}|$ so that $|\mathcal{C}|+1$ equals 1. Then the steady state satisfies the recursion:

$$q_k = \frac{R_k}{\alpha} J + \frac{1}{\alpha^2} q_{k+1}.$$

Fix $q_1 = q_{|\mathcal{C}|+1}$. Then:

$$
\begin{aligned}
q_{|\mathcal{C}|} &= \frac{R_{|\mathcal{C}|}}{\alpha} J + \frac{1}{\alpha^2} q_1 \\
q_{|\mathcal{C}|-1} &= \frac{R_{|\mathcal{C}|-1}}{\alpha} J + \frac{1}{\alpha^2} \left(\frac{R_{|\mathcal{C}|}}{\alpha} J + \frac{1}{\alpha^2} q_1 \right) \\
q_{|\mathcal{C}|-2} &= \frac{R_{|\mathcal{C}|-2}}{\alpha} J + \frac{1}{\alpha^2} \left(\frac{R_{|\mathcal{C}|-1}}{\alpha} J + \frac{1}{\alpha^2} \left(\frac{R_{|\mathcal{C}|}}{\alpha} J + \frac{1}{\alpha^2} q_1 \right) \right) \\
&\vdots \\
q_{|\mathcal{C}|-k} &= \sum_{j=0}^{k} \frac{R_{|\mathcal{C}|-j}}{\alpha^{2(k-j)+1}} J + \frac{1}{\alpha^{2(k+1)}} q_1
\end{aligned}
$$

Carrying the recursion all the way back to q_1:

$$
q_1 = q_{|\mathcal{C}|-(|\mathcal{C}|-1)} = \sum_{j=0}^{|\mathcal{C}|-1} \frac{R_{|\mathcal{C}|-j}}{\alpha^{2(|\mathcal{C}|-j)-1}} J + \frac{1}{\alpha^{2|\mathcal{C}|}} q_1
$$

Solving for q_1:

$$
q_1 = \frac{\alpha}{1-\alpha^{-2|\mathcal{C}|}} \sum_{j=0}^{|\mathcal{C}|-1} \left[\frac{R_{|\mathcal{C}|-j}}{\alpha^{2(|\mathcal{C}|-j)}} \right] J = \frac{\alpha}{1-\alpha^{-2|\mathcal{C}|}} \sum_{k=1}^{|\mathcal{C}|} \left[\frac{R_k}{\alpha^{2k}} \right] J
$$

The same analysis would work for any initial node. Therefore the steady state at any node can be written as a sum over the resistances around the loop, weighted by α^{2k} where k is the distance from a given edge to the node of interest, in the direction of rotation. To solve for the proper distribution we need to enforce normalization. This requires that the sum of the steady state probabilities is one. Since we have the steady state as a sum over all the edges this sum is a double sum over all nodes, and all edges. We can reverse the order of the sum so that we first perform a sum over all nodes given a fixed edge. This requires working out the contribution of a given edge to the steady state at each node. Given edge

k the contribution to node $k-1$ is $\propto \frac{R_k}{\alpha^2}$, to node $k-2$ is $\propto \frac{R_k}{\alpha^4}$. This is carried out all the way around the loop, so the contribution of a given edge is:

$$R_k J \frac{\alpha}{1-\alpha^{-2|\mathcal{C}|}} \sum_{j=1}^{|\mathcal{C}|} \alpha^{-2j} = R_k J \frac{\alpha}{1-\alpha^{-2|\mathcal{C}|}} \frac{1-\alpha^{-2|\mathcal{C}|}}{\alpha^2-1} = R_k J \frac{\alpha}{\alpha^2-1}.$$

Summing over all edges gives the normalization constant Z:

$$Z = J \frac{\alpha}{\alpha^2-1} \sum_k R_k = J \frac{\alpha}{\alpha^2-1} R$$

where $R = \sum_k R_k$ is the total resistance of the loop.

So, normalizing the expression for q_1:

$$q_1 = \frac{\alpha^2-1}{1-\alpha^{-2|\mathcal{C}|}} \sum_{k=1}^{|\mathcal{C}|} \left[\alpha^{-2k} \frac{R_k}{R} \right]. \tag{7.9}$$

Notice that:

$$\sum_{k=1}^{|\mathcal{C}|} \alpha^{-2k} = \frac{1-\alpha^{-2P}}{\alpha^2-1}$$

so Equation (7.9) can be interpreted as a weighted average of R_k/R around the loop, with weights set to the geometric distribution with parameter $\alpha^{-2} = \exp(-2\theta)$. Alternatively, the steady state is the convolution of the distribution of resistances with the exponentially decaying kernel $\exp(-2\theta)$. Since $\theta > 0$ this distribution is decaying as we consider edges farther and farther away from the node of interest. It follows that q_1 is a weighted average of R_k/R that is weighted more heavily towards edges near node 1 in the direction of rotation. This extends easily to different initial nodes. In all cases we perform a weighted average of R_k/R around the loop in the direction of rotation starting from the node of interest, and with weights set to the geometric distribution with parameter $\exp(-2\theta)$. An example is

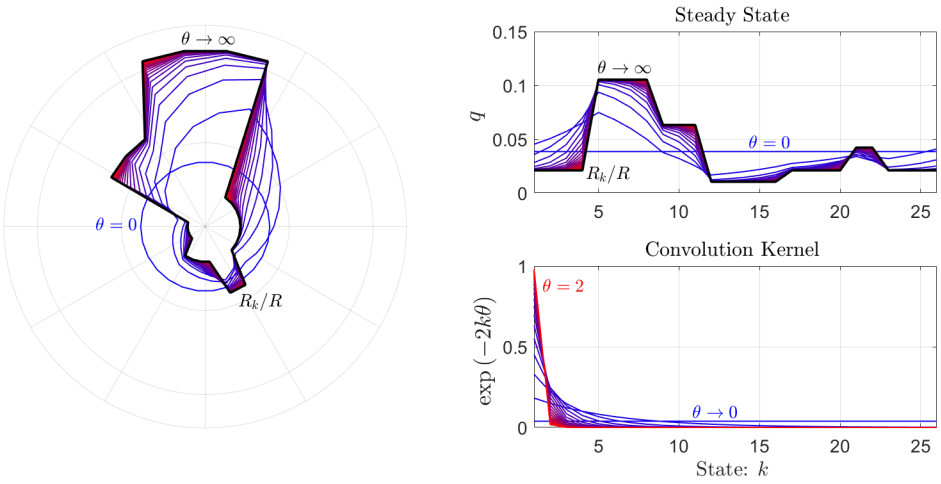

Figure 7.1: The steady state distribution q for a purely rotational process on a loop. The resistances on the loop are represented by the solid black lines. The steady state for different θ are represented by the blue and red lines. Blue lines correspond to θ near 0, and red lines correspond to θ near 2. In the strong rotation limit, $\theta \to \infty$ the steady state converges to the distribution of resistances. The bottom right panel shows the convolution kernel $\exp(-2k\theta)$.

illustrated in Figure 7.1

Equation (7.9) makes it easy to evaluate the weak rotation and strong rotation limits of the steady state. First, suppose rotation is strong. Then $\alpha = \exp(\theta)$ is very large. Then the geometric distribution decays very rapidly. It follows that the weighted average is weighted mostly towards the first edge leaving the node of interest. Therefore, in the limit as θ goes to infinity:

$$\lim_{\theta \to \infty} q_k(\theta) = \frac{R_{k,k+1}}{R}. \tag{7.10}$$

That is, in the strong rotation limit the steady state distribution converges to the distribution of resistances on the edges leaving each note. This result is entirely natural. In the strong rotation limit the flux across each edge is almost entirely due to probability flowing in the forward direction, so to force an equal current through all edges the probability in front of edges with a large resistance must be larger. Large resistances produce bottlenecks,

and probability accumulates on the lee side of edges with large resistances.

It is worth noting how different this result is from the limit as ϕ diverges. Then the distribution converges towards a delta distribution, and is dominated more and more by nodes with low potential. Therefore the potential directly determines the distribution. Here, in the large θ limit, the distribution does not converge towards a delta, but rather towards the distribution of resistances - which is entirely independent of θ.

On the other hand, if θ is very small then $\alpha \approx 1 + \theta$ so $\alpha^{-2k} \approx 1 - 2k\theta$. Then the weighted average converges to:

$$\lim_{\theta \to 0} q_1(\theta) \approx \frac{1}{|\mathcal{C}|(1 - (|\mathcal{C}|+1)\theta)} \sum_{k=1}^{|\mathcal{C}|} (1 - 2k\theta) \frac{R_k}{R} \approx 1 - 2\theta \sum_{k=1}^{|\mathcal{C}|} k \frac{R_k}{R}. \tag{7.11}$$

It follows that the steady state converges to a uniform distribution as rotation becomes weak.

This analysis provides an intuitive interpretation of the steady state of a purely rotational loop. The steady state distribution moves continuously from a uniform distribution to the distribution of resistances as the strength of rotation increases. When rotation is strong the skeleton process approaches a regular periodic cycle, so, in order to maintain a constant flux on all edges the probability of occupying a node immediately before a slow edge must increase proportional to the resistance on the edge. Large resistances lead to bottlenecks, where probability builds up until it is sufficient to match the steady state flux around the loop. If the direction of rotation is reversed then the exact same arguments apply in the opposite direction. As a result, if probability bottlenecks at the first endpoint of an edge under positive rotation, then it will bottleneck at the second endpoint under the reversed rotation. In contrast, when rotation is weak the process is dominated by diffusion, so approaches a uniform steady state distribution.

What is the steady state flux around the loop?

Since the flux is the same on all edges we need only evaluate it across one edge. On the first edge:

$$J = \frac{1}{R_{12}} \left(\alpha q_1 - \alpha^{-1} q_2 \right).$$

Both q_1 and q_2 are given by a sum over all edges. Consider the sequence of edges from 2 to 1 in the direction of rotation. Each of these edges contribute $\alpha^{-2k} R_k$ to q_1 and $\alpha^{-2(k-1)} R_k$ to q_2 (for now the other normalizing constants are ignored since they are shared). Then, multiplying the first by α and dividing the second by α we find that the first edge contributes $\alpha^{-2k+1} R_k$ to the flux from q_1 and $\alpha^{-2k+1} R_k$ to the flux from q_2. Therefore the difference between the two is zero. It follows that only the edge from 1 to 2 contributes to the flux. Therefore, replacing q_1 and q_2 with the contribution from the edge $1, 2$ to both:

$$J = \frac{1}{R_{12}} \frac{R_{12}}{R} \frac{\alpha^2 - 1}{1 - \alpha^{-2|\mathcal{C}|}} \left(\alpha \alpha^{-2} - \alpha^{-1} \alpha^{-2|\mathcal{C}|} \right) = \frac{1}{R} \frac{\alpha^2 - 1}{1 - \alpha^{-2|\mathcal{C}|}} \left(\alpha^{-1} - \alpha^{-2|\mathcal{C}|-1} \right).$$

Simplifying:

$$J = \frac{1}{R} \frac{\alpha^2 - 1}{1 - \alpha^{-2|\mathcal{C}|}} \frac{\left(1 - \alpha^{-2|\mathcal{C}|} \right)}{\alpha} = \frac{\alpha - \alpha^{-1}}{R}.$$

Finally, plugging in $\alpha = \exp(\theta)$:

$$J = 2 \frac{\sinh(\theta)}{R}. \tag{7.12}$$

Equation (7.12) is very natural. The hyperbolic sine is monotonically increasing so $\sinh(\theta)$ is effectively a measure of the strength of rotation. Therefore J can be interpreted as the net strength of rotation divided by the total resistance of the loop. When θ is large the steady state flux is large and when θ is small the steady state flux is small. On the other hand, when the resistance is large the flux is relatively small, and when the resistance

is small the flux is relatively large. Moreover, since $\sinh(\theta)$ is an odd function in θ, the steady state flux given $-\theta$ is equal to $-J$ given θ. That is, the magnitude of the steady state flux is independent of the sign of θ, and the direction matches the direction of θ.

Using Equation (7.12) and Equation (6.40) the steady state entropy production is:

$$P = 2f_{\text{rot}}^{\intercal} J = \frac{4|\mathcal{C}|\theta \sinh(\theta)}{R} = \frac{4\theta \sinh(\theta)}{\bar{R}} \tag{7.13}$$

where where R is the total resistance on the loop and $\bar{R}$ is the average resistance of the edges in the loop. Note that the total resistance R is the sum, $\sum_k R_k$, since the edges all appear in series.

In the weak rotation limit the fluxes converges to:

$$\lim_{\theta \to 0} J(\theta) \simeq \frac{2}{R}\theta. \tag{7.14}$$

The scaling between θ and CJ in the weak rotation limit is the Onsager coefficient [5]. Therefore, for a single loop, the Onsager coefficient is $2|\mathcal{C}|/R = 2/\bar{R}$ [1] .

Alternatively, when θ is large:

$$\lim_{\theta \to \infty} J(\theta) \simeq \frac{\exp(\theta)}{R}. \tag{7.15}$$

Therefore, for large driving forces the steady state current is exponentially large in the driving force. In the large θ limit the probability of a backward transition vanishes relative to the probability of observing a forward transition. The probability of observing a

[1] When we solve for the Onsager coefficients for arbitrary systems we will drop the factor of two from the definition of the coefficients since we can associate the factor of 2 with θ so that f_{rot} can be interpretted as work

backward or forward transition is:

$$p_- = \frac{\exp(-\theta)}{\frac{R_-}{R_+}\exp(\theta) + \exp(-\theta)}, p_+ = \frac{\exp(\theta)}{\exp(\theta) + \frac{R_+}{R_-}\exp(-\theta)}$$

where R_+ and R_- are the resistances on the forward and backward edges next to the node occupied by the process. Suppose for now that all the resistances are the same. Then p_- and p_+ only depend on θ and are constant across all the nodes.

It follows that the probability of observing exactly k consecutive forward reactions is $p_+^k p_-$. Therefore the length of streaks of forward reactions is geometrically distributed, and the expected number of forward reactions before a backward reaction is:

$$\mathbb{E}[\text{consecutive forward reactions}] = \frac{p_+}{p_-} = \exp(2\theta).$$

That is, the expected length of streaks of purely forward reactions is exponential in θ. Alternatively, the expected length of consecutive backward reactions is $\exp(-2\theta)$. Therefore, when θ is large we expect the random walk to consist of long streaks of forward reactions, with expected length exponential in θ, and that each streak of forward reactions is occasionally broken by a single backward reaction[2]. This result motivates consideration of the process with all backward transitions removed.

When all backward transitions are removed the process walks one node at a time in the forward direction around the loop. The steady state flux is one over the expected time to complete the loop. This expected time is the sum of the expected times to complete each step. The expected time to complete each step is one over the forward rate since the

[2]Note that a similar limiting argument will be used in Section 7.3.3 to consider the behavior of a Markov process driven by strong rotation on an arbitrary connected network

backward rates have been removed. Therefore:

$$J \simeq \left[\sum_k R_k \exp\left(-\theta\right)\right]^{-1} = \frac{\exp\left(\theta\right)}{R}.$$

which matches Equation (7.15).

Now that we have solved the problem for purely rotational processes we return to the original problem for arbitrary nonequilibrium processes. Suppose that the original process is not purely rotational. Then the fluxes of the purely rotational process match the fluxes of the original process. The steady states are not the same. The steady state of the original process is given by rescaling the steady state of the purely rotational process. In the strong rotation limit:

$$\lim_{\theta\to\infty} q_k(\theta) = \frac{1}{Z} q_k^{(eq)} \frac{R_k}{R} = \frac{1}{Z} \exp\left(-2\phi_k\right) \frac{\exp\left(\phi_k + \phi_{k+1}\right)}{\rho_{k,k+1} R} = \frac{1}{RZ} \frac{\exp\left(\phi_{k+1} - \phi_k\right)}{\rho_{k,k+1}}. \tag{7.16}$$

Therefore the steady state in the strong rotation limit is large at node k when the rotational forces point to $k+1$ and $\phi_{k+1} >> \phi_k$. This conclusion is natural since it means that probability accumulates at nodes where a large flux needs to be forced uphill against the scalar potential. It follows that the steady state in the strong rotation limit is large where the scalar potential changes quickly, not necessarily where it is large or small.

The associated effective potential (negative log of the steady state) is:

$$\phi_{eff_k}(\theta) = \log(-q_k(\theta)) = (\phi_{k+1} - \phi_k) - \log(\rho_{k,k+1}) \tag{7.17}$$

where $\rho_{k,k+1}$ is the conductivity on edge k. Therefore, in the strong rotation limit, the effective potential depends on the slope of the scalar potential, and the log conductances. The effective potential is small at nodes preceding a bottleneck. A bottleneck can be created

either by a large resistance (small conductance), or by a strong rotational flow pushing probability uphill against the conservative edge flow.

In summary, the steady state for an arbitrary system with isolated loops can be derived by considering the steady state for a purely rotational process on a single loop. This analysis shows that the strength of rotation acts as a shape parameter for the distribution, which is near uniform in weak rotation, and converges to the distribution of resistances in strong rotation. The steady state probability accumulates in nodes preceding a bottleneck. A bottleneck is a nodes with a large driving force into and out of the node with a large resistance on the edge leaving the node, or where the rotational forces point upwards against the scalar potential. In addition the steady state fluxes of the purely rotational system are proportional to $\sinh(\theta)/R$ where R is the total resistance of the loop, so are an odd function of the rotational potential, converge to a linear function of θ when θ is small, and diverge exponentially when $|\theta|$ is large.

7.2.3 Linked Loops

Section 7.2.2 provides a complete description of the nonequilibrium steady state and steady state fluxes of any nonequilibrium Markov process with isolated loops directly in terms of θ, ϕ and the resistances. Solving the same problem for more general network topologies is notoriously challenging. The steady state can be found directly by searching for the null space of the Laplacian, and the fluxes can be computed from the steady state, but this approach offers no real insight into how the conductances, scalar potential, and rotational forces combine to produce a steady state and steady state currents. It is impossible to ask how the steady states change if the conductances, scalar potential, or rotational forces are changed without recomputing the Laplacian, then the steady state. In stark contrast, if the process is an equilibrium process (obeys detailed balance), then the steady state only

depends on the scalar potential, can be computed directly from the potential using the Boltzmann equation eq. (6.10), and the steady state fluxes are all zero. Unfortunately this sort of simple description of nonequilibrium steady states is largely unavailable. Different authors have provided different characterizations of nonequilibrium steady states [257]. For example, Schnakenberg provides a characterization in terms of the work over an ensemble of optimal spanning trees [5]. The goal of this section is to illustrate the principal algebraic difficulty that makes finding a simple closed form for nonequilibrium steady states in terms of the conductances and edge flow close to impossible for general networks. These difficulties motivate the limiting approach taken in Section 7.3.

Consider the simplest network with a pair of linked loops - a pair of triangles sharing an edge (see Figure 3.3). As usual we can always scale out any equilibrium dynamics so that the process is purely rotational. So, without loss of generality, assume that the scalar potential is zero everywhere. Then the edge flow is determined entirely by the rotational potential on each loop. Let θ_I, θ_{II} denote the rotational potential on each triangle.

Suppose that $\theta_I = -\theta_{II} = \theta$. Then the edge flow on the edges in the outer loop all have the same magnitude, and the edge flow on the shared edge is 2θ. Orient the network so that the shared edge is vertical, and the edge flow points down the edge. Then the flow on the left triangle circulates clockwise, and the flow on the right triangle circulates counterclockwise. The two bottom edges of each triangle flow away from the shared edge, and the two top edges flow back into the shared edge (see Figure 7.2).

Assume that all the conductances are equal and set to one. On an isolated loop the steady state is uniform if the conductances are all the same and the scalar potential equals zero. Therefore it is reasonable to guess that the steady state should be uniform since there are no edges with large resistance to form bottlenecks. This assumption also corresponds to the intuition that the rotational forces should drive flux, not the accumulation of probability,

so a purely rotational system without any differing conductances should have a uniform steady state distribution.

Under the assumption that all the conductances are the same the transition rates in the forward direction of the edges in the boundary are all $\exp(\theta) = \alpha$, and the backward rates are $1/\alpha$. The forward rate on the shared edge is $\exp(2\theta) = \alpha^2$ and the backward rate is $1/\alpha^2$. Herin lies the obstacle. The forward transition rate along the shared edge is the square of the forward transition rates feeding into and out of the edge, so is greater than the sum of the pair of edges feeding into, and out of the edge. Thus, if $p = \mathbf{1}/4$ is uniform, the rate of probability flux over the shared edge is greater than the sum of the fluxes on the edges entering the shared edge, and the sum of the fluxes on the edges leaving the shared edge. As a result the flux has a nonzero divergence, and probability accumulates faster at the bottom node than it leaves. Therefore the steady state is not uniform.

Thus, the fact that the rotational flow associated with neighboring loops adds on their shared edges before exponentiation, but the flux leaving their shared edges is the sum of the exponential of the individual rotational flows, means that, starting from a uniform distribution, the flux generated on the shared edges is faster than the flux on the sum of the edges entering or leaving the shared edges, so the steady state is not uniform (see Figure 7.2). The nonlinearity of the exponential is the principal difficulty which makes finding a nonequilibrium steady state difficult. Even in the absence of large resistances to generate bottlenecks the steady state associated with a particular rotational potential is usually nonuniform, and depends on how the cycles overlap.

In the case of two linked loops it is possible to perform an analysis like the analysis performed in Section 7.2.2 by starting from the observation that the steady state current must be rotational, so only has two degrees of freedom. While tractable this analysis is not particularly insightful, requires much more work than the analysis for a single loop, and

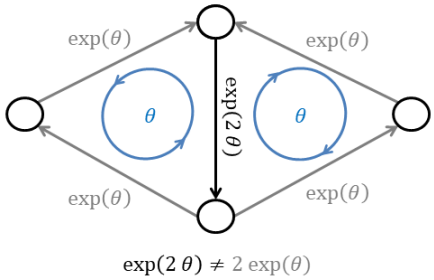

Figure 7.2: A pair of linked loops. The edge flow on the shared edge is twice the edge flow on any other edge, but the edge flow is exponentiated, so the net flux into and out of the shared edge does not equal the net flux across the edge if the distribution of probability is uniform.

is limited to pairs of linked loops. Therefore we adopt a different approach for studying nonequilibrium dynamics. In Section 7.3 we study nonequilibrium steady states and steady state fluxes in a variety of limits. These limits simplify the problem. When the rotational component, or overall flow, is small then the exponential can be linearized. When the rotational component, or overall flow, is large then timescale separation arguments can be used to simplify the problem.

7.3 Limiting Dynamics

In this section we consider the behavior of the steady state distribution and fluxes in a variety of limits. Throughout we consider transition rates parameterized by:

$$l(\beta)_{ij} = \rho_{ij}(\beta) \exp\left(\beta f_{ij}\right) \tag{7.18}$$

where β is analogous to an inverse temperature, $1/(k_B T)$ in a physical system, the conductances $\rho(\beta)_{ij} = \rho(\beta)_{ji}$ are the geometric average of the forward and backward rates, and the flow $f_{ij} = -f_{ji}$ is the log geometric difference in the forward and backward rates.

Limits are considered by making β large or small. The limits considered are:

1. *The weak forcing limit* (see Section 7.3.1), in which $\rho_{ij}(\beta)$ are assumed to be constant in β and β is small. This models systems whose edge flow is small, and whose forward and backward transition rates are close to symmetric. This models a system dominated by diffusion - so would apply to physical systems at high temperature, or to discretizations of stochastic differential equations whose diffusion term is larger than their drift term.

2. *The weak rotation limit* (see Section 7.3.2), in which f is purely rotational, $\rho_{ij}(\beta)$ are assumed to be constant in β, and β is small. This limit can be used to study the steady state of any nonequilibrium Markov process with a small rotational component if the scaling introduced in Section 7.2.1 is used to transform an arbitrary nonequilibrium process into a purely rotational process. Thus this limit generalizes the results of the weak forcing limit to allow for an arbitrarily large conservative component.

3. *The strong forcing limit* (see Section 7.3.3), in which $\rho_{ij}(\beta)$ are assumed to be constant in β and β is large. This models a system whose forward transition rates are much larger than its backward rates, where the average time spent in most states converges to zero, and where the skeleton process becomes close to deterministic.

4. *The strong rotation limit* (see Section 7.3.3), in which f is purely rotational and β is large. We consider cases where the conductances are fixed in β, and where the conductances change with β to ensure that the expected waiting time at each node converges to a constant. This limit can be used to study the steady state of any nonequilibrium Markov process with a large rotational component if the scaling introduced in Section 7.2.1 is used to transform an arbitrary nonequilibrium process

into a purely rotational process. Thus this limit generalizes the results of the strong forcing limit to allow for an arbitrarily small conservative component.

5. *The near deterministic limit* (see Section 7.3.3), in which β is large, and $\rho_{ij}(\beta)$ are chosen so that the expected waiting time in each node converges to a finite, nonzero constant as β goes to infinity. In this limit the skeleton process becomes close to deterministic, while the expected rate of motion remains finite. This is analogous to the coarse graining of a stochastic differential equation in the small noise limit.

In general when β is small the process is dominated by diffusion, while if β is large the transitions become highly directed so the process is dominated by drift. When the process is dominated by diffusion the HHD controls the shape of the steady state distribution. If the process is drift dominated then an alternative class of potentials based on evaluating the work over optimal paths is required. These are analogous to the quasipotential used to analyze SDE's in the small noise limit [23], however the actual form of the steady state depends on the assumed limiting behavior of the conductances which control the expected time spent in each state. Multiple strong forcing limits are considered since the strong forcing limit depends heavily on what is assumed about the conductances. This discussion lays the groundwork for a more general comparison of the HHD and quasipotential frameworks which is saved for future work.

7.3.1 Weak Forcing Limit

Suppose that the edge flow f scales with some small parameter β and the conductances are fixed. In a thermodynamic setting f_{ij} is the work w_{ij} to move from j to i divided by the temperature k_BT, so $\beta = \frac{1}{k_BT}$. Then small β is analogous to a high temperature limit.

Parameterize the transition rates:

$$l_{ji} = \rho_{ij} \exp(\beta f_{ij})$$

where $\rho_{ij} = \rho_{ji}$ are the conductances and $f_{ij} = -f_{ji}$ is the edge flow as defined in Section 6.4. Then when β is small the forward and backward rates on each edge pair are similar, so the process is dominated by diffusion rather than by drift. If $\beta = 0$ then L is symmetric. If L is symmetric then the corresponding master equation is purely diffusive, and, in a continuum limit, would converge to an SDE with no advective term.

When β is small:

$$\begin{aligned} l_{ji}(\beta) &= \rho_{ij}(1 + f_{ij}\beta + \mathcal{O}(\beta^2)) \\ l_{ii}(\beta) &= -\sum_{j \in \mathcal{N}_i} \rho_{ij}(1 + f_{ij}\beta + \mathcal{O}(\beta^2)). \end{aligned} \tag{7.19}$$

More generally let:

$$L(\beta) = \sum_{n=0} L^{(n)} \beta^n. \tag{7.20}$$

Then:

$$l_{ji}^{(n)} = \rho_{ij} \frac{(f_{ij})^n}{n!}, \quad l_{ii}^{(n)} = -\sum_{j \in \mathcal{N}_i} \rho_{ij} \frac{(f_{ij})^n}{n!}. \tag{7.21}$$

This expansion converges since the Taylor series for the exponential converges everywhere.

Let $q(\beta)$ denote the steady state distribution. Then expand $q(\beta)$:

$$q(\beta) = \sum_{n=0}^{\infty} q^{(n)} \beta^n. \tag{7.22}$$

Stationarity of the steady state requires $L(\beta)q(\beta) = 0$, so, matching orders of β:

$$\begin{aligned} L^{(0)}q^{(0)} &= 0 \\ L^{(0)}q^{(1)} + L^{(1)}q^{(0)} &= 0 \\ &\vdots \\ L^{(0)}q^{(n)} + \ldots L^{(n)}q^{(n)} &= 0. \end{aligned}$$

Rearranging each equation produces a recursive sequence of correction equations which define the correction to $q^{(n)}$ up to order n:

$$\begin{aligned} L^{(0)}q^{(0)} &= 0 \\ L^{(0)}q^{(1)} &= -L^{(1)}q^{(0)} \\ &\vdots \\ L^{(0)}q^{(n)} &= -\sum_{j=1}^{n} L^{(j)}q^{(n-j)}. \end{aligned} \tag{7.23}$$

The zeroeth order approximation to $L(\beta)$ is $L^{(0)} = -G^\intercal WG$ where W is a diagonal weight matrix with diagonal entries equal to the conductances on each edge. The first order correction, $L^{(1)}$, has ij entries $\rho_{ij}f_{ij}$. This matrix is antisymmetric since $\rho_{ij} = \rho_{ji}$ and $f_{ij} = -f_{ji}$. Its diagonal entries are the first order correction to $-\sum_{j\in\mathcal{N}_i} l_{ji}(\beta)$ which is the same as the sum of the first-order correction of the off-diagonal entries. The correction can be written $L^{(0)} = -G^\intercal WFH$ where G is the gradient, W is the weight matrix with edge weights equal to the conductances, F is the diagonal $E \times E$ matrix with diagonal entries equal to the edge flow on each edge, and H is the same as the gradient but with all nonzero entries set to $+1$.

To see that $G^\intercal WFH$ produces the correct matrix consider $L_{ij}^{(1)} = [G]_i^\intercal WF[H]_j$. This

is a weighted inner product between the i^{th} column of G and the j^{th} column of H since W and F are diagonal. These correspond to nodes i and j and have rows corresponding to the edges. Each column is zero everywhere except at edges that connect to the corresponding nodes. Thus, if $i \neq j$ the only shared nonzero entry corresponds to the edge k with endpoints i and j. If $i(k) = i$ and $j(k) = j$ then $G_{ki} = -1$ and $H_{kj} = 1$ so the weighted inner product is $-W_{kk}F_{kk} = -\rho_{ij}f_{ij}$. If $j(k) = i$ and $i(k) = i$ then $G_{ki} = 1$ and $H_{kj} = 1$ so the weighted inner product is $W_{kk}F = \rho_{ij}f_{ij}$. The columns all sum to zero since $\mathbf{1}^\intercal G^\intercal WFH = [G\mathbf{1}]^\intercal WFH = \mathbf{0}^\intercal WFH = \mathbf{0}^\intercal$. Therefore the diagonal entries must equal the sum of the off-diagonal entries. Therefore, $-G^\intercal WFH$ is a $V \times V$ matrix with off-diagonal entries $\rho_{ij}f_{ij}$ and diagonal entries equal to the (negative) sum of their respective columns. The first order correction $L^{(1)}$ has the same form, so:

$$\begin{aligned} L^{(0)} &= -G^\intercal WG \\ L^{(1)} &= -G^\intercal WFH \end{aligned} \tag{7.24}$$

and, substituting into the first two correction equations:

$$\begin{aligned} -G^\intercal WGq^{(0)} &= 0 \\ -G^\intercal WGq^{(1)} &= G^\intercal WFHq^{(0)}. \end{aligned} \tag{7.25}$$

The matrix $L^{(0)} = -G^\intercal WG$ is the Laplacian for a weighted simple random walk with weights ρ (see Section 7.2.1). Since it is symmetric the corresponding steady state is uniform. Let $q^{(0)} = \mathbf{1}/V$. Then $L^{(0)}q^{(0)} \propto -G^\intercal WG\mathbf{1} = 0$ since the gradient of any constant equals zero. Therefore:

$$q^{(0)} = \frac{1}{V}\mathbf{1}. \tag{7.26}$$

The fact that the steady state converges to a uniform distribution as β goes to zero is

natural since, when $\beta = 0$, the process is entirely diffusive.

Substitute the uniform distribution into the right hand side of Equation (7.25) to compute the equation defining the first order correction. Since each row of H has exactly two nonzero entries both equal to $+1$ the product $H\mathbf{1} = 2\mathbf{1}$. Then $FH\mathbf{1} = 2f$ where f is the edge flow. Then the first order correction equation is:

$$G^{\intercal}WG = -\frac{2}{V}G^{\intercal}Wf \tag{7.27}$$

which is, exactly, a weighted Poisson equation of the type covered by Theorem 11. In particular, suppose we defined a pair of generalized potentials ϕ, θ that are solutions to the weighted HHD:

$$-W^{\frac{1}{2}}G\tilde{\phi} + W^{-\frac{1}{2}}C^{\intercal}\tilde{\theta} = W^{\frac{1}{2}}f. \tag{7.28}$$

Then $q_1 = \frac{2}{V}\tilde{\phi}$. Alternatively we could have chosen the weighted HHD:

$$-G\hat{\phi} + W^{-1}C^{\intercal}\tilde{\theta} = f \tag{7.29}$$

or:

$$-WG\hat{\phi} + C^{\intercal}\tilde{\theta} = Wf \tag{7.30}$$

and we would still have:

$$q_1 = \frac{2}{V}\hat{\phi}. \tag{7.31}$$

The fact that we could pick multiple weightings is a natural consequence of the fact that we can freely multiply the equation from the left by an invertible weight without changing the decomposition. In general it is helpful to work with multiple versions since the symmetrized version, Equation (7.28), retains orthogonality but the asymmetric version,

Equation (7.29), will be easier to interpret (see Section 7.3.2).

The corresponding effective potential is defined:

$$\phi_{eff}(\beta) = -\frac{1}{\beta}\log\left(q(\beta)\right). \tag{7.32}$$

In the large noise limit:

$$\phi_{eff}(\beta) = -\lim_{\beta\to 0}\log\left(\frac{1}{V} + \frac{2}{V}\hat{\phi}\beta + \mathcal{O}(\beta^2)\right) = -2\hat{\phi} - \frac{1}{\beta}\log\left(V\right) + \mathcal{O}(\beta^2).$$

Therefore:

$$\lim_{\beta\to 0}\frac{1}{\beta}\phi_{eff}(\beta) = -\frac{2}{V}\hat{\phi} - \frac{1}{\beta}\log\left(V\right) + \mathcal{O}(\beta^2) \tag{7.33}$$

regardless the conductivies ρ_{ij}.

Therefore, in the weak forcing limit the steady state behaves like the steady state of a process that satisfies detailed balance (see Equation (6.10)). The steady state converges to a uniform distribution exponentially in β, but the deviations from the uniform distribution take a Boltzmann like form with potential $\hat{\phi}$ defined by the weighted HHD with forces equal to f and weighted by the conductances.

This result is easy to interpret using the path integral interpretation of the weighted HHD (see Theorem 12). The difference in $\hat{\phi}$ between two nodes is the average work to traverse an ensemble of paths between the two nodes against the edge flow f, where the paths are sampled from a simple random walk with weights W. This simple random walk is exactly the $\beta = 0$ limit of the original process. So when the forces are weak, the ensemble of paths used in the path integral interpretation of the scalar potential $\hat{\phi}$ is exactly the ensemble that would be sampled from the process in the limit. Thus it is not surprising that the weighted HHD describes the first order correction to the steady state in the weak forcing

limit. These observations are extended in Section 7.3.2 which considers the weak rotation limit at length. Since any Markov chain can be rescaled to produce a purely rotational process (see Section 7.2.1) the in-depth study of weak rotation will also cover all processes with weak forcing, but will allow for an arbitrarily large conservative component.

7.3.2 Weak Rotation

Suppose that the rotational component of the edge flow f_{rot} is small. Then the corresponding Markov process is near to a Markov process which obeys detailed balance. Since the equilibrium in detailed balance is well understood it is natural to seek a perturbative theory of steady states near detailed balance. That theory is developed here, using the same analytic approach developed for the weak forcing limit (Section 7.3.1) after using the purely rotational transform developed in Section 7.2.1.

Parameterize:

$$l_{ji}(\beta) = \rho_{ij} \exp(f_{\text{con}_{ij}} + \beta f_{\text{rot}_{ij}})$$

where β is a small parameter. Then scale the transition rates by $\exp(-2\phi)$ to produce the transformed process with rates:

$$\hat{l}_{ji}(\beta) = \frac{1}{R_{ij}} \exp(\beta f_{\text{rot}_{ij}}) \tag{7.34}$$

where the resistances $R_{ij} = R_{ji}$ are defined to be one over the rate of transition between nodes in the equivalent conservative process (see Section 7.2.1 and Section 6.3.1).

Let:

$$q_i^{\text{eq}} = \exp(-2\phi_i) \tag{7.35}$$

denote the equilibrium distribution corresponding to the original process without its rota-

tional component. Assume that ϕ is shifted so that $\exp(-2\phi_i)$ is normalized. Let $\hat{q}(\beta)$ denote the steady state for the process scaled by q^{eq}. Then:

$$q(\beta)_i = \frac{1}{Z} q_i^{\text{eq}} \hat{q}_i(\beta) \tag{7.36}$$

for the appropriate choice of the normalization constant Z.

Then, when β is small, the steady state of the scaled process $\hat{q}(\beta)$ is the steady state of a process in the weak forcing limit, so must satisfy the recursive sequence of correction equations:

$$\begin{aligned} \hat{L}^{(0)}\hat{q}^{(0)} &= 0 \\ \hat{L}^{(0)}\hat{q}^{(1)} &= -\hat{L}^{(1)}\hat{q}^{(0)} \\ &\vdots \\ \hat{L}^{(0)}\hat{q}^{(n)} &= -\sum_{j=1}^{n} \hat{L}^{(j)}\hat{q}^{(n-j)} \end{aligned} \tag{7.37}$$

where $\hat{L}(\beta) = \hat{L}^{(0)} + \hat{L}^{(1)}\beta + \hat{L}^{(2)}\beta^2 + \dots$. The columns of $\hat{L}^{(n)}$ sum to zero for any n, and the off-diagonal entries follow from the Taylor expansion of $\exp(f_{ij})$:

$$\hat{l}_{ji}^{(n)} = \frac{1}{R_{ij}} \frac{{f_{\text{rot}}}_{ij}^{n}}{n!}.$$

The key to the weak rotation expansion is to rewrite these matrices in terms of the gradient, resistances, and rotational edge flow. Let R be the diagonal $E \times E$ matrix with diagonal entries equal to the resistances, let F_{rot} be the diagonal $E \times E$ matrix with diagonal entries equal to the rotational component of the edge flow, f_{rot}, and let H be the $E \times V$ matrix which is zero in all entries where the gradient has zero entries, and one in all entries where the gradient is ± 1. Then, following Equation (7.24) we will attempt to expand $\hat{L}^{(n)}$

with a product of the form:

$$\hat{L}^{(n)} = -G^{\intercal} R^{-1} F_{\text{rot}} K^{(n)}$$

where $K^{(n)}$ is either G or H, as specified in Equation (7.38) below.

This expansion is motivated by the following two observations. First, any matrix of the form $G^{\intercal}A$ has columns which sum to zero since $\mathbf{1}^{\intercal}G^{\intercal}A = [G\mathbf{1}]^{\intercal}A$ and the gradient of any constant, i.e. $G\mathbf{1}$, equals zero. Second, any matrix of the form $G^{\intercal}WG$ or $G^{\intercal}WH$ where W is diagonal has the same sparsity pattern off the diagonal as the adjacency matrix of the original graph, and has nonzero off diagonal entries equal to $\pm w_{ij}$. If the matrix is of the form $G^{\intercal}WG$ then it is symmetric, if it is of the form $G^{\intercal}WH$ then its off-diagonal entries are antisymmetric. These two facts were shown in Section 7.3.1 and Section 7.2.1.

The matrices $\hat{L}^{(n)}$ are symmetric if n is even since $f_{\text{rot}_{ij}}^{\,n}$ is nonnegative if n is even. The off-diagonal entries of $\hat{L}^{(n)}$ are antisymmetric if n is odd since $f_{\text{rot}_{ij}} = -f_{\text{rot}_{ij}}$ and $f_{\text{rot}_{ij}}^{\,n}$ has the same sign as $f_{\text{rot}_{ij}}$. Therefore:

$$\hat{L}^{(n)} = -\frac{1}{n!} G^{\intercal} R^{-1} F_{\text{rot}}^{n} K^{(n)} \text{ where } K^{(n)} = \left\{ \begin{array}{l} G \text{ if } n \text{ even} \\ \\ H \text{ if } n \text{ odd} \end{array} \right\}. \tag{7.38}$$

Taken together Equation (7.37) and Equation (7.38) define a sequence of recursive correction equations:

$$\begin{aligned} -G^{\intercal} R^{-1} G \hat{q}^{(0)} &= 0 \\ -G^{\intercal} R^{-1} G \hat{q}^{(n)} &= -\sum_{j=1}^{n} -G^{\intercal} R^{-1} F_{\text{rot}} K^{(j)} \hat{q}^{(n-j)}. \end{aligned} \tag{7.39}$$

This sequence of equations is the weak rotation expansion. It can be solved one order at a time by solving a linear system. Once $\hat{q}^{(0)}$ is known then it is used on the right hand side

of the recursive equation in (7.39) to solve for $\hat{q}^{(1)}$, which is used on the right hand side to solve for $\hat{q}^{(2)}$, and so on. By solving this sequence of linear equations out to an order n we recover the coefficients in the n^{th} order Taylor expansion of $\hat{q}(\beta)$. Each term, $\hat{q}^{(n)}$ is the n^{th} derivative of $\hat{q}(\beta)$ at $\beta = 0$ divided by $n!$.

We will start by exploring the first order correction in Section 7.3.2, where we show that the first order correction term, $\hat{q}^{(1)}$ is the scalar potential associated with a weighted HHD. This result follows exactly the same logic as the weak forcing expansion, however we take the analysis further, and show that the first order correction to the steady state fluxes is associated with the rotational part of the weighted HHD via a change of weights of the type discussed in Corollary 11.1. This establishes a trade-off between the efficiency with which the small rotational component drives steady state fluxes, and the amount the steady state is perturbed. We show that the efficiency and size of perturbations depend on the variation in the resistances, since variation in the resistances produce bottlenecks, and that perturbations to the steady state arise from a balance between diffusion and bottlenecks. In Section 7.3.2 we show that there is a symmetric positive semi-definite linear mapping between the driving rotational component, f_{rot}, the steady state fluxes, and the steady state affinities. This mapping is the Onsager matrix. We provide an explicit formula for the Onsager coefficients, introduce bounds on the coefficients, and show that they are cycle basis independent. Finally we show that the first order steady state entropy production is minimized by the first order corrections to the steady state currents. To conclude our exploration of the first order-corrections we show that the space of possible fluxes generated by a process in detailed balance is limited to the range of a weighted gradient, and thus we can use the HHD to solve for a space of observables that are martingales. This observation is then extended in the weak rotation limit (see Section 7.3.2).

Once the first-order corrections are fully explained we return to the full expansion

Equation (7.39). We show that much of the interpretation developed for the first order corrections extends to all orders. In particular, at every order the correction to the steady state is the scalar potential associated with a weighted HHD, and the correction to the steady state fluxes at the matching order is the rotational component of the weighted HHD. This result proves that the HHD is fundamental to steady state dynamics of nonequilibrium processes near detailed balance, and governs the trade off between current and steady state at every order (see Section 7.3.2). In Section 7.3.2 we prove that the expansion always has a nonzero radius of convergence, and conjecture that the expansion converges as long as $\beta||f_{\text{rot}}||_\infty < 1$. The conjecture is inspired by numerical experiments on large percolation networks generated by removing nodes from a high dimensional hypercube. To conclude our analysis of the weak rotation limit we ask, is there any space of conductances/resistances such that the steady state is independent of rotation for any f_{rot}? We show that, if there are any cycles with overlapping edges, then there is no set of conductances for which the steady state is independent of rotation, however it is possible to choose a set of conductances such that the steady state is independent of rotation up to a particular order which depends on the number of degrees of freedom in f_{rot} (see Section 7.3.2). This result supports our observation in Section 7.2.3 that the principal difficulty when solving for the steady states of nonequilibrium processes is overlapping cycles and the nonlinearity of the exponential.

The First Order Corrections

Focus on the zeroeth and first order correction equations. These mimic the correction equations derived for weak forcing. The zeroeth order equation reads $G^{\intercal}WG\hat{q}^{(0)} = 0$ which is satisfied when $\hat{q}^{(0)}$ is uniform since the gradient of a constant is zero. Set $\hat{q}_i^{(0)} = \mathbf{1}$ so that $q^{(eq)}\hat{q}(\beta)$ is normalized for all β. Setting $\hat{q}^{(0)}$ to $\mathbf{1}/V$ would normalize $\hat{q}$ but this

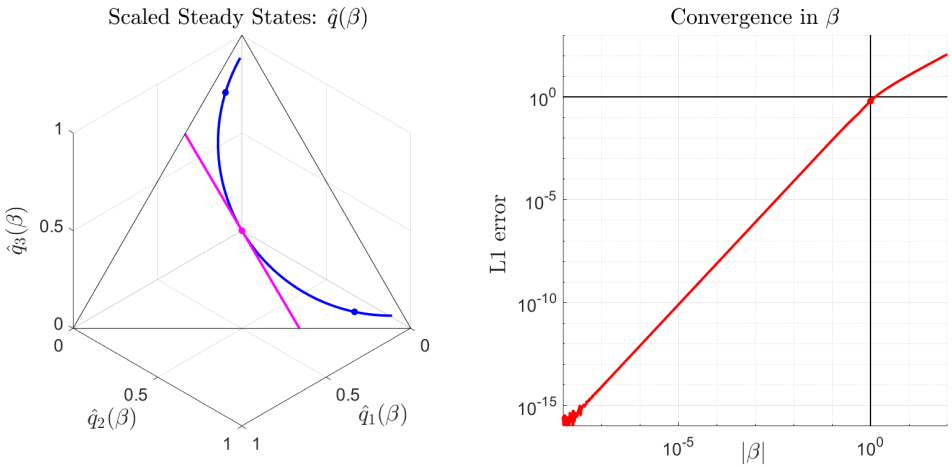

Figure 7.3: The first order approximation to $\hat{q}(\beta)$ using Equation (7.40) to compute $\hat{q}^{(1)}$ on a network with three nodes. The steady state probability $\hat{q}(\beta)$ of the transformed system is shown in blue. The two dots represent $\beta = \pm 1$. The magenta line is the first order approximation, and the magenta dot is detailed balance ($\beta = 0$). Note that the first order approximation is tangent to $\hat{q}(\beta)$ at detailed balance. Also note that, after using the purely rotational transform, the steady state at $\beta = 0$ is uniform. The right hand panel shows the L_1 error between the first order approximation and the scaled steady state. Note that the error decays with slope 2 since the first order approximation is accurate to order β^2.

scaling would ultimately vanish since $q^{(eq)}\hat{q}$ must be normalized.

Since $\hat{q}^{(0)} = \mathbf{1}$ the product $H\hat{q}^{(0)}$ is simply the $2\mathbf{1}$. Therefore the first order correction equation reads:

$$G^\intercal R^{-1} G\hat{q}^{(1)} = -G^\intercal R^{-1} F_{\text{rot}} H\hat{q}^{(0)} = -2G^\intercal R^{-1} f_{\text{rot}}. \tag{7.40}$$

This is a weighted Poisson equation, so is the solution to the weighted HHD with weights R^{-1} and right hand side $2f_{\text{rot}}$:

$$-G\hat{q}^{(1)} + RC^\intercal \hat{\theta}^{(1)} = 2f_{\text{rot}}. \tag{7.41}$$

The corresponding linear approximation to the steady state is shown in Figure 7.3

Let $J_{ij} = l_{ji}q_i - l_{ij}q_j = l_{ji}q_i^{(eq)}\hat{q}_i - l_{ij}q_j^{(eq)}\hat{q}_j = \hat{l}_{ji}\hat{q}_i - \hat{l}_{ij}\hat{q}_j$ be the steady state flux across the edge from j to i. If the process obeys detailed balance then this flux vanishes on every edge at equilibrium. When the process does not obey detailed balance then the rotational component of the edge flow typically leads to non-vanishing steady state fluxes. Consider an arbitrary set of fluxes J [257]. If the divergence of the fluxes at any node is nonzero then there is a net inflow or outflow of probability at that node. This is impossible at steady state, since a net inflow or outflow of probability changes the distribution. Therefore $G^\intercal J = 0$ at steady state. It follows immediately from the HHD that the steady state currents must be of the form $J = C^\intercal \theta_J$ for some $\theta_J \in \mathcal{R}^L$. Since the steady state and transition rates both depend on β these currents should be written $J(\beta) = C^\intercal \theta_J(\beta)$. Then the currents can be expanded in β. Since the currents vanish in detailed balance $J(0) = 0$. Therefore the currents have an expansion of the form $J(\beta) = J^{(1)}\beta + J^{(2)}\beta^2 + ...$

Expand the product $J_{ij}(\beta) = \hat{l}_{ji}(\beta)\hat{q}_i(\beta) - \hat{l}_{ij}(\beta)\hat{q}_j(\beta)$ to order zero in β:

$$J_{ij}(\beta) = \hat{l}_{ji}^{(0)}\hat{q}_i^{(0)} - \hat{l}_{ij}^{(0)}\hat{q}_j^{(0)} + \mathcal{O}(\beta).$$

The zeroeth order term, $\hat{l}_{ji}^{(0)}\hat{q}_i^{(0)} - \hat{l}_{ij}^{(0)}\hat{q}_j^{(0)} = 0$ since the zeroeth order process obeys detailed balance, so generates no flux at equilibrium. This leaves the first order term. The first order term can be written:

$$J(\beta) = \left[R^{-1}F_{\text{rot}}H\hat{q}^{(0)} + R^{-1}G\hat{q}^{(1)}\right]\beta + \mathcal{O}(\beta^2)$$

where $R^{-1}F_{\text{rot}}H$ is the first order term in the matrix that maps from probability to fluxes, and $R^{-1}G$ is the zeroeth order term in the matrix that maps from probability to fluxes. These relations come from removing the divergence $-G^\intercal$ from $\hat{L}^{(1)}$ and $\hat{L}^{(2)}$. The Laplacians without the divergence are the mapping from probability to flux since the Laplacians

map from probability to rate of change in probability, and the rate of change in probability equals the divergence of the fluxes. Move the factor of R^{-1} to the outside. The product $H\hat{q}^{(0)} = \mathbf{2}$ since $\hat{q}^{(0)} = \mathbf{1}$. Then:

$$J(\beta) = R^{-1}\left[2f_{\text{rot}} + G\hat{q}^{(1)}\right]\beta + \mathcal{O}(\beta^2) \tag{7.42}$$

There are two natural ways to interpret this expression. First, notice that the two terms in square brackets can each be associated with different sets of edge flows. The first set are the small rotational forces applied to the system, f_{rot}, responsible for preventing the system from reaching steady state at $\hat{q}^{(0)}$. The second arise from the fact that the stationary distribution is near the equilibrium distribution $\hat{q}^{(0)}$ but not at $\hat{q}^{(0)}$. Therefore $\hat{L}^{(0)}$ drives fluxes that move the system back towards $\hat{q}^{(0)}$.

Alternatively, note that the first order term, $J^{(1)}$, is the rotational component left over in the correction HHD (Equation (7.41)) so:

$$-G\hat{q}^{(1)} + RC^{\intercal}\hat{\theta}^{(1)} = -G\hat{q}^{(1)} + RJ^{(1)} = 2f_{\text{rot}} \tag{7.43}$$

where $J^{(1)} = C^{\intercal}\hat{\theta}^{(1)}$ is the first order approximation to the steady state currents. Thus, the first order corrections to the steady state and first order corrections to the steady state fluxes are the conservative and rotational components of the same weighted HHD, with weights set to the resistances. Therefore the HHD governs the first order steady state dynamics of any Markov process near to detailed balance.

Lemma 26 (First order corrections in weak rotation)**.** *In the weak rotation limit the first order corrections to the (scaled) steady state, $\hat{q}^{(1)}$, and steady state fluxes, $J^{(1)}$, satisfy the*

weighted HHD:

$$-G^{\intercal}\hat{q}^{(1)} + RC^{\intercal}\theta_J = 2f_{rot}$$

where $J^{(1)} = C^{\intercal}\theta_J$ and R is the diagonal matrix with entries equal to the resistances.

Equation (7.43) shows that there is an inherent trade-off in the size of the steady state fluxes and the amount that the equilibrium distribution is perturbed by rotational forces. If the fluxes are particularly large then the steady state does not respond significantly to the rotational forces, but if the fluxes are small then the steady state is highly responsive to the rotational forces. This exchange can be made more precise by symmetrizing the correction HHD. Multiply both sides of the equation by $R^{1/2}$. Then:

$$-R^{1/2}G\hat{q}^{(1)} + R^{-1/2}C^{\intercal}\hat{\theta}^{(1)} = 2R^{1/2}f_{\text{rot}}.$$

The scaled operators, $R^{1/2}G$ and $R^{-1/2}C^{\intercal}$, are orthogonal since:

$$(R^{1/2}G)^{\intercal}R^{-1/2}C = G^{\intercal}R^{1/2}R^{-1/2}C^{\intercal} = G^{\intercal}C^{\intercal} = 0.$$

Therefore $R^{1/2}G\hat{q}^{(1)}$ and $R^{-1/2}C^{\intercal}\hat{\theta}^{(1)} = R^{-1/2}J^{(k)}$ are orthogonal. Then, by the Pythagorean theorem:

$$||G\hat{q}^{(1)}||^2_{R^{-1}} + ||J^{(1)}||^2_R = 4||f_{\text{rot}}||^2_{R^{-1}} \tag{7.44}$$

where $||v||^2_A = v^{\intercal}Av$ denotes the energy norm in A for a symmetric positive definite matrix A.

The right hand side is fixed by the rotational component of the edge flow and resistances, so is independent of the corrections at order 1. Since both terms on the left hand side are non-negative both are bounded above by the right hand side, and if one is close in

magnitude to the right hand side then the other must be small. Therefore the more of the decomposition is devoted to fluxes the smaller the necessary correction to the steady state, and visa versa.

This trade-off leads to the hypothesis that the more efficiently a set of rotational forces drive steady state currents, the less the steady state will deviate from the equilibrium distribution. The efficiency with which the rotational forces produce current is measured by the ratio:

$$\eta^2 = \frac{||J^{(1)}||_R^2}{4||f_{\text{rot}}||_{R^{-1}}^2}. \tag{7.45}$$

To see that η is an efficiency note that the steady state currents arise from the rotational forces. The numerator measures the size of the currents, and the denominator measures the size of the forces, so η is a natural measure of efficiency. Going beyond heuristic arguments, $\eta > 0$ since the numerator is nonnegative and the denominator is positive by assumption. The ratio η is strictly less than or equal to one since since $||G\hat{q}^{(1)}||_W^2 + ||J^{(1)}||_{W^{-1}}^2 = 4||f_{\text{rot}}||_W^2$ and the square of the energy norm of $G\tilde{q}^{(1)}$ is nonnegative. Moreover $\eta = 1$ can be achieved since if the resistances are uniform ($R_{ij} = r$ for all ij). If the resistances are all the same, then $G^\intercal R^{-1} f_{\text{rot}} = rG^\intercal C^\intercal \theta = 0$ so $\hat{q}^{(1)} = 0$. Then $\eta = 1$. Therefore the network is most efficient at producing rotation when the resistances are uniform, and when the network is maximally efficient, rotation does not change the steady state. Finally, η is dimensionless and compares the actual entropy production of the network to the maximal entropy production that could be achieved. Entropy will be addressed in more depth in Section 7.3.2. Here we briefly discuss entropy to motivate the otherwise strange fact that in the numerator we use an energy norm evaluated with respect to R and in the denominator we use an energy norm with respect to R^{-1}.

Consider an analogy with an electrical network. On a given edge J is a probability

current. It would have units of charge per time. On any edge R^{-1} is analogous to a resistance, where R would have units of Ohms. The edge flow f_{rot} would be analogous to a change in energy across the edge [5], so would have units of volts, V. Then:

$$||J||_R^2 = \sum_{ij} R_{ij} J_{ij}^2.$$

is the Joule heating of the electrical network. In Section 7.3.2 we will show that this is also the entropy production of the Markov chain. Therefore it has units of power. On the other hand:

$$||f_{\text{rot}}||_{R^{-1}}^2 = \sum_{ij} \frac{V_{ij}^2}{R_{ij}}.$$

Note that in order for f_{rot} to be rotational V should be interpreted as the voltage drop across a battery on each edge. If the network was an electric circuit, then, when the voltages and currents are small it would obey Ohms law, $V = IR$, so this would also be the Joule heating and would have units of power. Therefore η is a ratio of the actual entropy production of the network to the maximal possible entropy production, which is achieved when the network behaves like an electric circuit and obeys Ohms law. Thus η is the thermal efficiency of the network in the weak rotation limit.

The correction to the steady state is related to the thermal efficiency by:

$$||G\hat{q}^{(1)}||_{R^{-1}}^2 = 4\left(1 - \eta^2\right) ||f_{\text{rot}}||_{R^{-1}}^2. \tag{7.46}$$

Therefore, when the network is maximally efficient the correction to the steady state vanishes. The less efficiently the rotational forces drive current the larger the change to the steady state. Let $\sigma_{min}, \sigma_{max}$ be the smallest and largest singular values of the weighted node Laplacian, $G^{\intercal} R^{-1} G$. Then the size of the correction to the steady state is bounded

by:

$$4\frac{1-\eta^2}{\sigma_{max}}||f_{\text{rot}}||^2_{R^{-1}} \leq ||\hat{q}^{(1)}||^2 \leq 4\frac{1-\eta^2}{\sigma_{min}}||f_{\text{rot}}||^2_{R^{-1}}. \tag{7.47}$$

Equation (7.47) shows that the steady state is more reactive to the rotational forces when the rotational forces do not efficiently drive current. Conversely, when the steady state is reactive to the rotational forces the currents must be produced inefficiently. To understand this exchange it is helpful to return to the weighted Poisson equation defining the correction to the steady state:

$$G^{\intercal}R^{-1}G\hat{q}^{(1)} = -2G^{\intercal}R^{-1}f_{\text{rot}}.$$

First consider the product $R^{-1}f_{\text{rot}}$. The rotational flow f_{rot} are unitless while the resistances R have units of time. Therefore $-2R^{-1}f_{\text{rot}}$ has units one over time, the same units as probability flux. If there is no correction to the steady state the correction HHD gives $J^{(1)} = 2R^{-1}f_{\text{rot}}$, therefore $2R^{-1}f_{\text{rot}}$ can be interpreted as the maximal probability fluxes. Note that, from the expansion of the fluxes, the maximal fluxes are the fluxes which arise from applying $L^{(1)}$ to the zeroeth order steady state $\hat{q}^{(0)}$.

The maximal fluxes need not be divergence free. When they are not divergence free the actual fluxes are smaller since the rotational flow does not drive current with perfect efficiency. Suppose the maximal fluxes are not divergence free. Then at some nodes there will be more maximal flux entering the node than leaving it, and at others there will be less maximal flux entering the node than leaving it. This pulls probability away from nodes where the divergence of $R^{-1}f_{\text{rot}}$ is positive and towards nodes where the divergence of $R^{-1}f_{\text{rot}}$ is negative. Probability bottlenecks where the divergence is negative because more probability is pushed in than is removed. Therefore $-2G^{\intercal}R^{-1}f_{\text{rot}}$ is the rate at which probability would accumulate in each node under the maximal fluxes alone. The rate is large when there are bottlenecks: nodes where more flux enter than leaves. These

bottlenecks couple the steady state distribution to the rotational component f_{rot}.

The accumulation of probability in bottlenecks is balanced by the inherently diffusive nature of random processes. The diffusion is captured by the left hand side of the discrete Poisson equation, and corresponds to the flux generated by the product of $L^{(0)}$ with $\hat{q}^{(1)}$. The more probability builds up in a given node the more it diffuses to its neighbors. The left hand side of the Poisson equation is the rate of change in probability due to diffusion. Therefore the Poisson equation can be interpreted as the steady state equation for a diffusive process with sources $2G^{\intercal}R^{-1}f_{\text{rot}}$. The singular values that appeared in Equation (7.47) represent the maximal possible rate and minimal possible rate at which diffusion spreads probability between the nodes. Since the flow of probability due to diffusion increases as $\hat{q}^{(1)}$ becomes farther from uniform it eventually balances the accumulation of probability in bottlenecks. It is this balance that determines the steady state. The maximal fluxes $R^{-1}f_{\text{rot}}$ are not always divergence free, leading to bottlenecks that accumulate probability. The accumulation is balanced by the inherently diffusive nature of random walks, leading to a steady state that is large near nodes where the maximal fluxes bottleneck, and small where $G^{\intercal}R^{-1}f_{\text{rot}}$ is negative.

It follows that the size of the correction to the steady state should depend on both the size of the optimal fluxes, and their tendency to bottleneck. To characterize the tendency to bottleneck we seek a bottlenecking coefficient, which, when multiplied with the magnitude of the maximal currents, characterizes the size of the correction to the steady state.

Define the bottlenecking coefficient, γ:

$$\gamma^2 = \frac{||G^{\intercal}R^{-1}f_{\text{rot}}||^2_{(G^{\intercal}R^{-1}G)^{\dagger}}}{4||f_{\text{rot}}||^2_{R^{-1}}}. \tag{7.48}$$

where $(G^{\intercal}R^{-1}G)^{\dagger}$ is the pseudo-inverse of the weighted node Laplacian, $\hat{L}^{(0)}$, responsible

for diffusion of probability. The choice to use an energy norm in the pseudo-inverse of $G^\intercal R^{-1} G$ may seem strange. The choice is natural since bottlenecks lead to larger changes in the steady state when they occur at nodes which probability diffuses from slowly. Like the thermal efficiency the bottlenecking coefficient is a dimensionless quantity. It is independent of both the average size of R and f_{rot} since both appear in the numerator and denominator to the same order. It is large when the $G^\intercal R^{-1} f_{\text{rot}}$ is uncharacteristically large compared to $R^{-1} f_{\text{rot}}$. This occurs when the maximal fluxes tend to converge or diverge. That is, when the maximal fluxes bottleneck. It is small when the maximal fluxes are close to divergence free, and vanishes when the maximal fluxes are divergence free. This last property is essential since, when the maximal fluxes are divergence free they do not bottleneck, so the coefficient should be zero.

The bottlenecking coefficient, is intimately related to the efficiency, η, with which the f_{rot} produces current. In fact:

Lemma 27 (Bottlenecking and efficiency)**.** *The bottlenecking coefficient, γ, defined by Equation (7.48) and the thermal efficiency, η, defined by Equation (7.45) satisfy:*

$$\gamma^2 + \eta^2 = 1. \tag{7.49}$$

Proof. To prove Equation (7.49) expand the energy norm in the numerator of the bottlenecking coefficient:

$$||G^\intercal R^{-1} f_{\text{rot}}||^2_{(G^\intercal R^{-1} G)^\dagger} = \frac{1}{4}||G^\intercal R^{-1} G\hat{q}^{(1)}||^2_{(G^\intercal R^{-1} G)^\dagger} = \frac{1}{4}||G\hat{q}^{(1)}||^2_{R^{-1}} = (1-\eta^2)||f_{\text{rot}}||^2_{R^{-1}}.$$

Then $\gamma^2 = 1 - \eta^2$, $\gamma^2 + \eta^2 = 1$, and $\eta^2 = 1 - \gamma^2$. □

Lemma 27 shows that there is an exact exchange between the degree to which f_{rot}

introduces bottlenecks and the efficiency with which f_{rot} drives current. The more bottlenecking the less efficiently current is driven, and the more efficiently current is driven the less bottlenecking. Equation (7.49) also shows that, like the thermal efficiency, the bottlenecking coefficient is bounded above by one. The bottlenecking coefficient equals one when the rotational forces fail to produce any current (to first order).

The size of the first order correction steady state can then be bounded by the bottlenecking coefficient by substituting Equation (7.49) into Equation (7.47):

$$4\frac{\gamma^2}{\sigma_{max}}||f_{\text{rot}}||^2_{R^{-1}} \leq ||\hat{q}^{(1)}||^2_{R^{-1}} \leq 4\frac{\gamma^2}{\sigma_{min}}||f_{\text{rot}}||^2_{R^{-1}}. \tag{7.50}$$

Thus the larger the bottlenecking coefficient, the larger the change to the steady state distribution due to rotation.

Suppose now that the resistances are all constants so that $R = rI$. Then there are no bottlenecks since $G^\intercal R^{-1} f_{\text{rot}} = r^{-1} G^T f_{\text{rot}} = 0$. If $G^\intercal R^{-1} f_{\text{rot}} = 0$ then the bottlenecking coefficient is zero, so the first order correction to the steady state is zero, and the thermal efficiency is one. Therefore, if $R_{ij} = r$ for all connected pairs of nodes ij then $\hat{q}^{(1)} = 0$, $\beta = 0$, $\eta = 1$, and $J^{(1)} = \frac{1}{r} f_{\text{rot}}$ on every edge. In Section 7.3.2 we will show that when the resistances are all equal then the nonequilibrium steady state is independent of rotation to third order in β ($\hat{q}^{(2)} = 0$), but at third order and higher still depends on rotation. In Section 7.3.2 we study conditions for rotation independent steady states and find that the coupling to rotation at higher orders arises from to the nonlinearity of the exponential where loops overlap discussed in Section 7.2.3.

Onsager and Linear Thermodynamics

Linear thermodynamics is the study of nonequilibrium physical processes that are in the weak rotation limit (near to equilibrium) [257, 258]. In this section we show that some of the key results of linear thermodynamics regarding fluxes and entropy production can be recovered directly from the first order terms in the weak rotation expansion.

Consider the first order approximation to the steady state currents, $J^{(1)}$. The first order approximation to the steady state currents are the rotational component of the weighted correction HHD $-G\hat{q}^{(1)} + RJ^{(1)} = 2f_{\text{rot}}$ where $f_{\text{rot}} = C^{\intercal}\theta$. Take the curl on both sides. Then, since the curl of the gradient is zero:

$$CRJ^{(1)} = 2Cf_{\text{rot}} \tag{7.51}$$

or:

$$CRC^{\intercal}\theta_J = 2CC^{\intercal}\theta \tag{7.52}$$

where $C^{\intercal}\theta_J = J^{(1)}$. It is always possible to find a rotational potential so that $C^{\intercal}\theta_J = J^{(1)}$ since $J^{(1)}$ are steady state currents, so must be divergence free. Alternatively, the currents are necessarily divergence free since $J^{(1)} = R^{-1}[2f_{\text{rot}} + G\hat{q}^{(1)}]$ so $G^{\intercal}J^{(1)} = 2G^{\intercal}R^{-1}f_{\text{rot}} + G^{\intercal}R^{-1}G\hat{q}^{(1)}$ which equals zero by the discrete Poisson equation (Equation (7.40)). Therefore the vorticities θ_J describing the steady state current are related to the rotational potential θ by the change of weights formula used to switch from an unweighted to a weighted HHD.

Equation (7.52) defines a mapping from the rotational component of the edge flow to the steady state currents:

$$J^{(1)} = C^{\intercal}\theta_J = 2C^{\intercal}[CRC^{\intercal}]^{-1}CC^{\intercal}\theta = 2C^{\intercal}[CRC^{\intercal}]^{-1}Cf_{\text{rot}}. \tag{7.53}$$

This relation can be extended to understand the steady state production rate of observables. Suppose $S[X(t)]$ is an observable that only changes when $X(t)$ transitions, where the change associated with a transition depends only on the transition made, and is reversed if the transition is made in reverse. Then $S[X(t)]$ can be expressed as a path integral over the trajectory $X(t)$ against an edge flow s and the long term production rate of $S[X(t)]$ is:

$$\lim_{t\to\infty} \frac{1}{t} S[X(t)] = J^{\intercal} s \tag{7.54}$$

where J is the steady state flux. Then, since J must be divergence free, $J = C^{\intercal}\theta_J$ for some vorticity θ_J, so $J^{\intercal}s = \theta_J^{\intercal} C s$. Since s is an edge flow $s = s_{\text{con}} + s_{\text{rot}}$ so $Cs = Cs_{\text{rot}}$ and $\theta_J^{\intercal} C s = J^{\intercal} s_{\text{rot}} = \theta_J^{\intercal} C C^{\intercal} \theta_s$. Then the long term production rate of the observable is:

$$\lim_{t\to\infty} \frac{1}{t} S[X(t)] = s_{\text{rot}}^{\intercal} J = \theta_s^{\intercal}[CJ]. \tag{7.55}$$

Therefore, the long term production rate of any observable is determined by the curl of steady state fluxes. Like the flux itself, the curl of the steady state flux is related to the driving rotational edge flow and θ by a linear relation in the weak rotation limit:

$$CJ^{(1)} = CC^{\intercal}\theta_J = 2CC^{\intercal}[CRC^{\intercal}]^{-1}CC^{\intercal}\theta = 2M\theta. \tag{7.56}$$

Thus the long term production rate of any observable defined by a path integral is expressable as an inner product with $CC^{\intercal}[CRC^{\intercal}]^{-1}CC^{\intercal}\theta$. The matrix, M, is the matrix of Onsager coefficients. The Onsager matrix maps from the driving rotational potential to the net flux on each cycle [5]. A similar matrix, $2C^{\intercal}[CRC^{\intercal}]^{-1}C$ can be introduced which maps from the driving rotational edge flow to the flux on each edge. These two matrices are central to thermodynamics near detailed balance since they govern the relationship between

the external affinities which drive rotation, and the change in observables due to rotating probability current.

Consider the steady state affinities, $A_{ij}(q)$ in the weak rotation limit. The steady state affinities are the change in free energy required to move an infinitesimal amount of probability along any edge. In general the affinity on edge ij given the distribution p is defined (see Chapter 6):

$$A_{ij}(p) = \log\left(\frac{l(\beta)_{ij}p_j}{l(\beta)_{ji}p_i}\right).$$

Therefore, at steady state:

$$A_{ij}(q|\beta) = \log\left(\frac{l(\beta)_{ij}q(\beta)_j}{l(\beta)_{ji}q(\beta)_i}\right)$$

Expanding in small β yields:

$$A_{ij}(q|\beta) = \log\left(\frac{l^{(0)}_{ij}q^{(0)}_j + \beta[l^{(0)}_{ij}q^{(1)}{}_j + l^{(1)}_{ij}q^{(0)}{}_j] + \mathcal{O}(\beta^2)}{l^{(0)}_{ji}q^{(0)}{}_i + \beta[l^{(0)}_{ji}q^{(1)}{}_i + l^{(1)}_{ji}q^{(0)}{}_i] + \mathcal{O}(\beta^2)}\right).$$

To simplify note that $l^{(0)}_{ij}q^{(0)}_j = R^{-1}_{ij} = l^{(0)}_{ji}q^{(0)}_i$. Therefore multiplying the numerator and denominator by R_{ij} gives:

$$A_{ij}(q|\epsilon) = \log\left(\frac{1 + \beta R_{ij}[l^{(0)}_{ij}q^{(1)}{}_j + l^{(1)}_{ij}q^{(0)}{}_j] + \mathcal{O}(\beta^2)}{1 + \beta R_{ij}[l^{(0)}_{ji}q^{(1)}{}_i + l^{(1)}_{ji}q^{(0)}{}_i] + \mathcal{O}(\beta^2)}\right).$$

Separating into a difference of logarithms and Taylor expanding each logarithm in small β gives:

$$A_{ij}(q|\epsilon) = \beta R_{ij}\left[l^{(0)}_{ij}q^{(1)}{}_j - l^{(0)}_{ji}q^{(1)}{}_i + l^{(1)}_{ij}q^{(0)}{}_j - l^{(1)}_{ji}q^{(0)}{}_i\right] + \mathcal{O}(\beta^2).$$

The bracketed term is $J_{ij}^{(1)}$ so:

$$A(q|\beta) = \beta R J^{(1)} + \mathcal{O}(\beta^2). \tag{7.57}$$

and:

$$A^{(1)}(q) = R J^{(1)} \tag{7.58}$$

Therefore the steady state affinities are all proportional to the steady state currents, and the steady state currents are the affinities divided by the resistances [5]:

$$\beta J^{(1)} = R^{-1} A^{(1)}(q) \tag{7.59}$$

This linear relationship between the fluxes and affinities is one of the key results of linear thermodynamics. It follows immediately that the curl of the steady state affinities is:

$$A_{\text{ext}}^{(1)} = C A^{(1)}(q) = C R J^{(1)} = 2CC^{\intercal}\theta.$$

which corresponds exactly to the general rule:

$$CA = C\left(2f + \log\left(\frac{p_j}{p_i}\right)\right) = 2Cf = 2CC^{\intercal}\theta.$$

The entropy production P is the inner product $P^{\intercal}J$ [20] so, to lowest order:

$$P^{(2)} = ||A^{(1)}||_{R^{-1}}^2 = ||J^{(1)}||_R^2 \tag{7.60}$$

Therefore the steady state entropy production is, to lowest order in β, the energy norm of the steady state affinities with respect to R^{-1}, or the energy norm of the steady state

fluxes with respect to the resistances R. Note that the latter energy norm was the numerator in the thermal efficiency η^2 (see Equation (7.45)). This fact motivates the name thermal efficiency, since η^2 is the ratio of the entropy produced to the maximum possible entropy produced.

Lemma 28 (Affinities, Flux, and Entropy Production in Weak Rotation)**.** *In the weak rotation limit the first order correction to the steady state fluxes and affinities are proportional where $J^{(1)} = R^{-1}A^{(1)}$, so to lowest order the steady state entropy production is:*

$$P(q|\beta) = \beta^2 ||J^{(1)}||_R^2 + \mathcal{O}(\beta^4). \tag{7.61}$$

Lemma 28 is interesting since $J^{(1)}$ is the circulating component of the weighted correction HHD, with weights R. Then the least squares interpretation of the weighted HHD (see Section 2.4.2) implies that $||J^{(1)}||_R^2$ is minimized by the first order correction to the steady state. Thus, in the weak rotation limit, the steady state and steady state fluxes minimize the entropy production. Other maximum power, maximum efficiency, and minimum dissipation properties of the weak rotation limit are discussed in [259].

The entropy production is an observable, and at steady state is defined as an inner product with the steady state fluxes. It follows that the entropy production can be computing directly from f_{rot} or θ using the Onsager matrix M.

$$\begin{aligned} P^{(1)} &= A^{(1)\mathsf{T}} J^{(1)} = A^{(1)\mathsf{T}} C^{\mathsf{T}} \theta_J = A_{\text{ext}}^{\mathsf{T}} \theta_J = 2\theta^{\mathsf{T}} C C^{\mathsf{T}} \theta_J \\ &= 2\theta^{\mathsf{T}} C J^{(1)} = 2\theta^{\mathsf{T}} M \theta^{\mathsf{T}} = 2||\theta||_M^2. \end{aligned}$$

That is, to first order the entropy production is twice the energy norm of the rotational potential with respect to the Onsager matrix M.

The Onsager matrix is the mapping from θ to the cycle fluxes $CJ^{(1)}$. This mapping

equals:

$$M = CC^{\intercal}[CRC^{\intercal}]^{-1}CC^{\intercal}. \tag{7.62}$$

The Onsager matrix can be simplified by expressing the product $C^{\intercal}[CRC^{\intercal}]^{-1}C$ as a projector. Symmetrize the correction HHD by multiplying on the left by $R^{-1/2}$. Then:

$$-R^{-1/2}G\hat{q}^{(1)} + R^{1/2}J^{(1)} = 2R^{-1/2}f_{\text{rot}}$$

where $J^{(1)} = C^{\intercal}\theta_J$. Let $\tilde{G} = R^{-1/2}G$ and $\tilde{C} = CR^{1/2}$. Then the weighted operators are orthogonal so $R^{1/2}J^{(1)}$ equals the orthogonal projection of $2R^{-1/2}f_{\text{rot}}$ onto the range of $R^{1/2}C^{\intercal}$. Let $P_{\tilde{C}}$ denote the orthogonal projector onto the range of $R^{1/2}C^{\intercal}$. Then $R^{1/2}J^{(1)} = 2P_{\tilde{C}}R^{-1/2}f_{\text{rot}} = 2P_{\tilde{C}}R^{-1/2}C^{\intercal}\theta$. Therefore:

$$\begin{aligned} J^{(1)} &= 2R^{-1/2}P_{\tilde{C}}R^{-1/2}f_{\text{rot}} \\ CJ^{(1)} &= 2CR^{-1/2}P_{\tilde{C}}R^{-1/2}C^{\intercal}\theta = 2M\theta \end{aligned} \tag{7.63}$$

and:

$$M = CC^{\intercal}[CRC^{\intercal}]^{-1}CC^{\intercal} = CR^{-1/2}P_{\tilde{C}}R^{-1/2}C^{\intercal}. \tag{7.64}$$

The coefficients of the Onsager matrix, m_{lh}, represent the coupling between the rotational potential θ on loop h, and the observed cycle current on loop l. If m_{lh} is large then introducing a rotational potential on loop h will lead to a large current around loop h. This formula for the Onsager matrix differs from Schnakenberg's formula for the Onsager matrix since Schnakenburg uses a curl operator which evaluates the curl of a flow on a fundamental cycle basis by evaluating the flow on each chord, rather than summing the flow around the each cycle [5].

For example, consider a single loop with $|\mathcal{C}|$ vertices. Then C is the $1 \times |\mathcal{C}|$ row vector

of all ones. Then $CRC^\intercal = \sum_{ij} R_{ij}$ is the total resistance of the cycle. Then $[CRC^\intercal]^{-1}$ is one over the total resistance of the loop. At the same time $CC^\intercal = |\mathcal{C}|$ so the Onsager coefficient for the scaled system is $|\mathcal{C}|/\bar{R}$ where $\bar{R}$ is the total resistance. This result differs from the result derived in Section 7.2.2 because the result derived there assumed that the initial process was purely rotational, while in this section we assumed that we were working with a nonequilibrium process that is scaled to arrive a purely rotational process. Since we wanted to avoid introducing a normalization constant Z we set $q^{(0)} = \mathbf{1}$ instead of $\mathbf{1}/V = \mathbf{1}/|\mathcal{C}|$. This distinction introduces a factor of $|\mathcal{C}|$ since the steady state of a purely rotational process on a cycle in the limit that θ goes to zero is $\mathbf{1}/V = \mathbf{1}/|\mathcal{C}|$.

The Onsager coefficients satisfy a number of interesting properties. These are summarized below:

Lemma 29 (Onsager Matrix)**.** *Let M be the matrix of Onsager coefficients defined by Equation (7.64). Then M depends only on equilibrium ($\beta = 0$) quantities and has the following properties:*

1. ***Symmetry:*** *M is symmetric so $m_{lh} = m_{hl}$ for all loops h and l.*

2. ***Positive Semi-Definite:*** *M is positive semi-definite so:*

$$\begin{aligned} m_{ll} &\geq 0 \\ m_{hl} &\leq \frac{1}{2}(m_{hh} + m_{ll}) \\ m_{hl} &\leq \sqrt{m_{hh} m_{ll}}. \end{aligned} \tag{7.65}$$

3. ***Bounded:*** *$||M||_2 \leq ||CR^{-1}C^\intercal||_2$ where $||A||_2$ denotes the induced two-norm. The upper bound scales in the inverse of the average resistance.*

4. ***Cycle Basis Independent:*** *Each Onsager coefficient depends only on the pair of cycles it maps between, and does not depend on any of the other cycles included in the set of cycles that defines the curl. If $\mathcal{C}$ and $\mathcal{C}'$ are two different cycle sets, sharing a pair of cycles, then the Onsager coefficient corresponding to that pair of cycles is the same for both $\mathcal{C}$ and $\mathcal{C}'$, and is well defined for any set of cycles whether or not they form a cycle basis.*

Proof. By construction M only depends on the network topology and the resistances R, which equal the rate at which probability is exchanged between neighbors at equilibrium.

The symmetry of the Onsager matrix M is clear from Equation (7.64):

$$M^\intercal = \left(CR^{-1/2}P_{\tilde{C}}R^{-1/2}C^\intercal\right)^\intercal = CR^{-1/2}P_{\tilde{C}}^\intercal R^{-1/2}C^\intercal$$

since R is diagonal. The projector $P_{\tilde{C}}$ is an orthogonal projector, so it is symmetric, thus $M^\intercal = M$.

The Onsager matrix is positive semi-definite since:

$$v^\intercal Mv = (R^{-1/2}C^\intercal v)^\intercal P_{\tilde{C}}(R^{-1/2}C^\intercal v) = u^\intercal P_{\tilde{C}}u \geq 0$$

where the inequality follows from the fact that the projector is positive semi-definite. Since M is positive semi-definite the inner product $(CJ^{(1)})^\intercal\theta = \theta^\intercal M\theta \geq 0$, so the steady state fluxes generated by θ tend to point in the same direction around the basis loops as θ.

Let $\mathcal{C}_l$ be a loop in the cycle basis. Then $m_{ll} = e_l^\intercal Me_l \geq 0$ where e_l is the l^{th} column of a $L \times L$ identity matrix. It follows that all of the diagonal entries of M are nonnegative. Thus, driving rotation around a loop with a rotational potential θ drives current in the same direction. The coupling between the rotational potential on a loop and the cycle flux on that

same loop is a diagonal coefficient of M.

The fact that M is positive semi-definite also bounds the off-diagonal entries of M given the diagonal entries. Let $\mathcal{C}_l$ and $\mathcal{C}_h$ be two different cycles in $\mathcal{C}$ then let $v = e_l - e_h$. Then $v^\intercal M v = m_{hh} + m_{ll} - 2m_{hl} \geq 0$ so:

$$m_{hl} \leq \frac{1}{2}(m_{hh} + m_{ll}).$$

That is, the off-diagonal entries of the Onsager matrix are less than or equal to the average of the corresponding pair of diagonal entries. This means that the coupling between the rotational potential on loop l, and the current induced by that potential on loop h is always less than the average of the coupling between loop l and itself and loop h and itself. Therefore any flux introduced by coupling between distinct loops is weaker than the flux induced on the loops themselves.

The same inequality is true if a geometric average is used instead of an arithmetic average. Write M as a block matrix of the form:

$$M = \begin{bmatrix} m_{11} & \bar{m_1}^\intercal \\ \bar{m_1} & \bar{M} \end{bmatrix}.$$

Then define:

$$T = \begin{bmatrix} 1 & 0 \\ -\bar{m_1}/m_{11} & I \end{bmatrix}.$$

Then:

$$TMT^\intercal = \begin{bmatrix} m_{11} & 0 \\ 0 & \bar{M} - \frac{1}{m_{11}}\bar{m}_1\bar{m}_1^\intercal \end{bmatrix}.$$

Let $v = T^{\intercal} e_k$ with $k > 1$. Then:

$$v^{\intercal} M v = m_{kk} - \frac{m_{1k}^2}{m_{11}} \geq 0.$$

It follows that $m_{1k}^2 \leq m_{11} m_{kk}$. Since the choice of blocking was arbitrary:

$$m_{jk} \leq \sqrt{m_{jj} m_{kk}}.$$

Therefore the off-diagonal elements of M are less than or equal to the arithmetic and geometric averages of the corresponding diagonal elements.

The norm of the Onsager matrix is bounded since:

$$\begin{aligned} ||M||_2 &= ||CR^{-1/2} P_{\tilde{C}} R^{-1/2} C^{\intercal}||_2 \leq ||CR^{-1/2}||_2 ||P_{\tilde{C}}||_2 ||R^{-1/2} C^{\intercal}||_2 \\ &= ||CR^{-1/2}||_2 ||R^{-1/2} C^{\intercal}||_2 = ||CR^{-1/2}||_2^2 = ||CR^{-1} C^{\intercal}||_2 \end{aligned}$$

where the first inequality follows from the definition of the induced two-norm, the second equality follows from the fact that the largest singular value of a orthogonal projector equals one, and the last two equalities follow from the fact that the largest singular value of a matrix equals the largest singular value of its transpose, and the singular values of the product of a matrix and its transpose are the singular values of the original matrix squared. If the resistances are scaled by a constant factor then $||CR^{-1}C^{\intercal}||_2$ is scaled by one over the same factor. Thus, if $\bar{r}$ denotes the average resistance, and $\delta r_{ij} = (r_{ij} - \bar{r})/\bar{r}$ is the deviation in the resistances relative to their mean, then if $\bar{r}$ changes while δr stays fixed then $||CR^{-1}C^{\intercal}||_2 \propto \bar{r}^{-1}$. Therefore the Onsager matrix is bound above by a bound that scales in one over the average resistance. Naturally, increasing the resistances decreases the loop currents.

To show that the entries of the Onsager matrix only depend on the corresponding pair of loops, and not on any other loops included or excluded from $\mathcal{C}$ when defining the curl C consider:

$$m_{lh} = [CR^{-1/2}P_{\tilde{C}}R^{-1/2}C^{\intercal}]_{lh} = C_l[R^{-1/2}P_{\tilde{C}}R^{-1/2}]C_h^{\intercal}$$

where C_l is the l^{th} row of the curl. The l^{th} and h^{th} row of the curl only depend on loops $\mathcal{C}_l$ and $\mathcal{C}_h$, and are independent of all other loops in $\mathcal{C}$. The bracketed term is entirely cycle basis independent since the resistances R don't depend on the cycle basis, and the orthogonal projector $P_{\tilde{C}}$ equals the identity minus the orthogonal projector onto the range of the scaled gradient $\tilde{G}$, which only depends on the resistances not the cycle basis. Therefore m_{lh} only depends on $\mathcal{C}_l$ and $\mathcal{C}_h$, not the set of cycles.

□

The conclusions of Theorem 29 are worth interpreting. The symmetry of the Onsager coefficients is the most famous and most surprising conclusion. The symmetry of the Onsager coefficients is called Onsager reciprocity [5, 260, 261]. Onsager won the 1968 Nobel Prize for Chemistry for his seminal work on reciprocity. As innocent as reciprocity may appear in algebra, m_{lh} and m_{hl} have totally different interpretations. The first is the current induced on loop l by a unit rotational potential on loop h. The second is the current induced on loop h by a unit rotational potential on loop l. Thus reciprocity states that the current induced on loop l by driving rotation on loop h is the same as the current induced on loop h by driving rotation on loop l, no matter how much the loops may differ in size, net resistance, or significance. If one loop represents a cycle in one state variable or observable, and the other represents a cycle in a different state variable or observable, then reciprocity implies that driving rotation in the first state variable leads to the same cyclic flux in the second state variable as driving rotation in the second state variable leads to flux in the

first. Thus the flux of totally different physical quantities is reciprocal near equilibrium [260, 261]. Examples even include the exchange of information and energy [262].

For an ecological example, suppose the Markov process models multiple populations interacting in multiple geographic patches. Then there may be some cycles associated with individuals dispersing between habitat patches, and some cycles that represent changes in the number of individuals in each patch due to internal demographic processes (births and deaths). Then a rotational potential in dispersal will lead to cyclic fluxes in population sizes within patches, and an equivalently sized rotational potential in demographic rates within a patch will result in the same flux in dispersal between patches.

The fact that the Onsager matrix is positive semi-definite is intuitive since it implies that, when a rotational potential is introduced, the resulting loop currents circulate, on average, in the same direction as the driving potential. Moreover it ensures that if rotation is introduced on a loop, then the resulting current on that loop must go in the same direction. This result is a weaker version of Hill's cycle flux theorem [6], which guarantees that the direction of flux around any loop is the same as the direction around the loop in which the work to complete one cycle is positive, and which was guaranteed by Equation (6.39). The fact that the off-diagonal entries of the Onsager matrix are bound above by the arithemtic and geometric averages of the corresponding diagonal entries is also natural since it implies that the flux produced on loop l by a rotational potential on loop l will necessarily be larger than the flux produced on a different loop h if the flux produced on loop h by driving current on loop h is less than the flux produced on loop l by driving current on l.

The upper bound on the norm of the Onsager matrix is important since it shows that less current is induced when the resistances increase.

The fact that the Onsager coefficient coupling a pair of loops is independent of all other loops in the cycle basis, or set of cycles used to define the curl, is essential since the choice

of cycle basis, or possible larger set of cycles, is an arbitrary choice of representation. If the Onsager coefficients depended on the entire set of cycles chosen then they would be properties of how we choose to represent the system, not inherent properties of the loops themselves. The Onsager coefficients are inherent properties of pairs of loops since they do not depend on the inclusion or exclusion of other loops in the cycle set.

In fact, since $P_{\tilde{C}} = I - P_{\tilde{G}}$ where $P_{\tilde{G}}$ is the orthogonal projector onto the range of $\tilde{G} = R^{-1/2}G$, the Onsager coefficients for a pair of loops can be computed without ever forming a cycle basis or a curl. Let $\mathcal{C}_l$ and $\mathcal{C}_h$ be a pair of loops. Let $C(\mathcal{C})$ be the curl associated with set of loops $\mathcal{C}$. Then the corresponding coefficient is:

$$M(\mathcal{C}_l, \mathcal{C}_h) = C(\mathcal{C}_l)[R^{-1/2}(I - P_{\tilde{G}})R^{-1/2}]C(\mathcal{C}_h)^\intercal. \tag{7.66}$$

Equation (7.66) can be used to compute the Onsager coefficients for a cycle formed by the sum of two cycles. Suppose cycles $\mathcal{C}_l$ and $\mathcal{C}_h$ share a boundary. Then:

$$\begin{aligned} M(\mathcal{C}_l + \mathcal{C}_h, \mathcal{C}_k) &= [C(\mathcal{C}_l) \pm C(\mathcal{C}_l)][R^{-1/2}(I - P_{\tilde{G}})R^{-1/2}]C(\mathcal{C}_h)^\intercal \\ &= M(\mathcal{C}_l, \mathcal{C}_k) \pm M(\mathcal{C}_h, \mathcal{C}_k) \end{aligned} \tag{7.67}$$

where the two are subtracted if $\mathcal{C}_l$ and $\mathcal{C}_h$ cross their shared boundary in the same direction, and added otherwise. Thus, if the matrix M is computed for a cycle basis, then the Onsager coefficients for any other pair of loops can be computed by linear combination of the entries of M.

It follows immediately that, given two cycle bases $\mathcal{C}, \mathcal{C}'$ whose curls are related by the linear transform T such that $C(\mathcal{C}') = TC(\mathcal{C})$, then the Onsager matrix on the transformed cycle basis is $M(\mathcal{C}')$:

$$M(\mathcal{C}') = TM(\mathcal{C})T^\intercal. \tag{7.68}$$

In summary, by using the HHD to study nonequilibrium processes near to detailed balance we can easily recover a number of crucial results from linear thermodynamics. The fluxes and affinities are proportional, the entropy production at steady state is minimized (under constraints on the steady state flux), and the steady state loop currents are related to the driving rotation by a symmetric positive definite matrix whose entries are inherent to the pairs of loops considered.

Observables and Fluxes in Detailed Balance and Weak Rotation

In the Section 7.2.1 we developed a new way to write the Laplacian for processes that obey detailed balance. Given L that obeys detailed balance, the resistances are defined $1/R_{ij} = \rho_{ij} \exp(-[\phi_i + \phi_j])$. Let R^{-1} be an $E \times E$ diagonal matrix with entries set to one over the resistances. Then define a diagonal matrix Q with diagonal entries set to $\exp(-2\phi) = q$ where q is the stationary distribution for the process. Then the transition matrix L could be rewritten $L = -[G^\intercal R^{-1} G]Q^{-1} = G^\intercal[-R^{-1}GQ^{-1}]$. where $[-R^{-1}GQ^{-1}]$ is the matrix that maps from probability to the probability flux.

Given $L = G^\intercal[-R^{-1}GQ^{-1}]$, $\frac{d}{dt}p = G^\intercal R^{-1}GUp = -G^\intercal J(p)$ where $J(p)$ are the fluxes across the undirected edges. This linear mapping follows immediately from the definition of R, G and Q (see Section 6.3.1):

$$[R^{-1}GQ^{-1}p]_{ij} = \frac{1}{R_{ij}}[p_i/q_i - p_j/q_j] = \rho_{ij}\exp(-[\phi_i+\phi_j])[\exp(2\phi_i)p_i - \exp(2\phi_j)p_j]$$
$$= w_{ij}[\exp(\phi_i - \phi_j)p_i - \exp(\phi_j - \phi_i)p_j] = l_{ji}p_i - l_{ij}p_j = J_{ij}(p).$$

Therefore, the fluxes $J(p)$ are always in the range of the operator $-R^{-1}GQ^{-1}$. Both of the scaling matrices R^{-1} and Q^{-1} are diagonal and invertible, so the range of $-R^{-1}GQ^{-1}$ is $R^{-1}\text{range}\{G\}$. Notice that the range of G always has dimension $V - 1$. Therefore the

fluxes are restricted to a $V-1$ dimensional subspace of the space of all possible fluxes. In contrast the space of all possible fluxes has dimension $E-1$ since probability must be conserved. Therefore the space of fluxes that could be generated by a process that obeys detailed balance is a subspace of the space of all possible fluxes if $E > V$. It also follows that the fluxes are always orthogonal to the range of $RC^{\intercal}$. That is $CRJ(p) = 0$ for any p. Note that CRJ evaluates the curl of the fluxes weighted by the resistance on each edge, so this is equivalent to Kirchoff's second law, which requires that the voltage drop across all the edges is curl free. This is a strong restriction on the fluxes of a system that obeys detailed balance.

The fact that the fluxes are always orthogonal to CR implies that there exists a space of observables that are conserved by processes that obey detailed balance. Define an observable S to be a function on trajectories $S[X]$ such that $S[x_1, x_2, ...x_n] - S[x_1, x_2, ...x_{n-1}]$ is given by an antisymmetric function s defined on the edges. That is, if $x_n = i$ and $x_{n-1} = j$ then $S[x_1, x_2, x_3, ...x_n] = S[x_1, x_2, x_3, ...x_{n-1}] + s_{ij}$ and $s_{ij} = -s_{ji}$. Then S is an action functional evaluated over paths. The work over trajectories is a familiar example. Notice that this definition also extends naturally to observables whose logarithm is a path integral. This occurs when the update to the observable after a given transition is a product of the original value of the observable with some non-negative function on the undirected edges, and the update when the reverse transition occurs is given by dividing by the same non-negative function.

Let $S(t) = S[X(t)]$. Then $S(t)$ is a random variable, and is a transform of the trajectory $X(t)$. The stochastic process $S(t)$ is a martingale (conserved in a probabilistic sense) if $\mathbb{E}[S(t+\Delta t)] = S(t)$ for any $\Delta t \geq 0$ [244]. Martingales are observables that are conserved in expectation.

The rate of change in the observable is given by:

$$\frac{d}{dt}\mathbb{E}[S(t)] = \sum_{ij} s_{ij} J_{ij}(t) = s^{\intercal} J(t) \tag{7.69}$$

where $J_{ij}(t)$ is the probability flux over the edge ij. Therefore, if $X(t)$ obeys detailed balance, then the inner product $s^{\intercal}J(t) = 0$ if $s \in \text{range}\{RC^{\intercal}\}$. Thus range$\{RC^{\intercal}\}$ defines a space of martingales.

Lemma 30 (The Space of Martingales in Detailed Balance)**.** *Given $X(t)$ governed by Laplacian L which obeys detailed balance any realization $S(t)$ of an observable $S[X]$ where $S[X]$ is a path integral over the trajectory X against s is a martingale (conserved in expectation) if:*

$$s \in \text{range}\{RC^{\intercal}\} = \text{null}\{R^{-1}G\}. \tag{7.70}$$

where R^{-1} is the diagonal weight matrix with diagonal entries equal to the rate of transition across each pair of directed edges at equilibrium of L.

For any S, s is an edge flow so there exists a pair of potentials ϕ_s, θ_s such that:

$$s = -G\phi_s + RC^{\intercal}\theta_s. \tag{7.71}$$

Then, by the linearity of the path integral, any action functional $S[X]$ defined as a path integral over an edge flow can be decomposed into $S_{\text{con}}[X]$ and $S_{\text{rot}}[X]$ such that:

$$S[X] = S_{\text{con}}[X] + S_{\text{rot}}[X] \tag{7.72}$$

where $S_{\text{con}}[X]$ is defined by evaluating path integrals over $-G\phi_s$ while $S_{\text{rot}}[X]$ is defined by evaluating path integrals over $RC^{\intercal}\theta_s$. By construction the rotational action $S_{\text{rot}}[X]$ defines

a martingale. Therefore:

$$\frac{d}{dt}\mathbb{E}[S(t)] = \frac{d}{dt}\mathbb{E}[S_{\text{con}}(t)]. \tag{7.73}$$

Any action functional defined by a conservative field $s_{\text{con}} = -G\phi_s$ obeys $S(t) = \phi_s(X(0)) - \phi_s(X(t))$. Therefore:

$$\frac{d}{dt}\mathbb{E}[S(t)] = \frac{d}{dt}\phi_s(X(t)). \tag{7.74}$$

And, more strongly:

$$\mathbb{E}[S(t)] = \phi_s(X(0)) - \mathbb{E}[\phi_s(X(t))]. \tag{7.75}$$

Therefore, if $p(t)$ approaches an equilibrium q in the long time limit then any observable of the form $S(t)$ evaluated over trajectories drawn from a process obeying detailed balance satisfy $\lim_{t\to\infty}\mathbb{E}[S(t)] = \phi_s(X(0)) - \mathbb{E}_q[\phi_s(X)]$ where the expected value is evaluated over the equilibrium q. This also leads to a stronger version of Lemma 30:

Theorem 31 (Observables in Detailed Balance)**.** *Suppose $X(t)$ is governed by Laplacian L which obeys detailed balance, and $S(t) = S[X(t)]$ where $S[X(t)]$ is a path integral over the trajectory X against the edge flow s. Then there exist unique potentials (up to the addition of a constant) ϕ_s, θ_s such that:*

$$s = -G\phi_s + RC^{\intercal}\theta_s$$

and the observed trajectory $S(t)$ obeys:

$$\mathbb{E}[S(t+h)] = \mathbb{E}[\phi_s(X(t))] - \mathbb{E}[\phi_s(X(t+h))].$$

Therefore $S(t)$ is a martingale if and only if $\phi_s(X(t))$ is a martingale. If $X(t)$ is a martingale for any initial condition then ϕ_s is constant, so, without loss of generality, is zero everywhere.

Proof. The fact that $S(t)$ is a martingale if and only if $\phi_s(X(t))$ is a martingale was proved above. If $X(t)$ could start at any node, and is a martingale for any initial condition then it must be a martingale for $X(0) = x_j$ for any $j \in \mathcal{V}$. Then $\mathbb{E}[\phi_s(X(0))] = \phi_s(x_j)$. Since the Markov chain has a unique steady state distribution approached from any initial condition $\mathbb{E}[\phi_s(X(t))]$ must approach the expected value of $\phi_s(X)$ where X is drawn from the steady state. It follows that if the process is a martingale for any initial condition then $\phi_s(x_j)$ equals the expected value of $\phi_s(X)$ when X is drawn from q for all possible j. Then $\phi_s(x_j) = \mathbb{E}_q[\phi_s(X)] = \phi_s(x_i)$ for all pairs of node ij, so the potential is constant. If the potential is constant then $G\phi_s = 0$ so $S_{\text{con}} = 0$, so, without loss of generality, ϕ_s is zero everywhere.

□

Now suppose that L does not obey detailed balance, but is in the weak rotation limit. Then the results discussed above can be extended to weak rotation by introducing order β perturbations to the subspace of possible fluxes.

In the weak rotation limit the fluxes are given by removing the divergence from $L^{(0)}$ and $L^{(1)}$:

$$J(p, \epsilon) = R^{-1} \left[-G + \beta F_{\text{rot}} H\right] Q^{-1} p + \mathcal{O}(\beta^2). \tag{7.76}$$

As before, our goal is to find the subspace of possible fluxes. Since the fluxes are related linearly to the probability distribution, which has dimension V, they are always restricted to, at most, a V dimensional subspace. Notice that in detailed balance this subspace was

$V-1$ dimensional. The difference in dimension was a result of the special symmetry of detailed balance. In general the fluxes are restricted to the range of whatever $E \times V$ linear operator maps from probability to fluxes. In detailed balance this operator is $-R^{-1}GQ^{-1}$, which has a one dimensional nullspace corresponding to the equilibrium distribution, so has rank $V-1$. Near detailed balance this operator is, to first order, $R^{-1}[-G+\beta F_{\text{rot}}H]Q^{-1}$, which may have rank V.

Consider a simple example. Suppose the network consists of three nodes arranged in a triangle, with the edges oriented clockwise around the triangle. Then let $\theta = 1$ and $\beta = 1/2$ so $f_{\text{rot}} = [1/2, 1/2, 1/2]$. Then:

$$-G+\beta F_{\text{rot}}H = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3 & -1 & 0 \\ 0 & 3 & -1 \\ -1 & 0 & 3 \end{bmatrix}$$

which has rank $3 = V$. Therefore, outside of detailed balance, the space of fluxes may have dimension V not $V-1$.

Since the linear operator mapping to fluxes from probability always has a range with dimension at most V the subspace of fluxes only spans the space of possible edge flows when $E \leq V$. If $E \leq V$ and the network is connected then either the network is a tree with $E = V-1$ or the network contains one cycle. Otherwise there is always always an $E - V = L - 1$ dimensional subspace orthogonal to the subspace of fluxes. As in detailed balance, this subspace defines a vector space of martingales. Notice that this subspace has dimension greater than zero if and only if the network has at least two loops, or has one loop and $F_{\text{rot}}H$ maps into the range of G. We start by studying the perturbation under the assumption that the space of martingales has dimension L instead of $L-1$. Typically the resulting subspace will have one extra dimension that does not correspond to a martingale.

By including the extra dimension we can develop a perturbation theory for a subspace which encloses the space of martingales. We then show that the extra dimension introduces a contradiction if it is assumed that all edge flows in the subspace produce martingales, identify the extra dimension, and discard it to produce the true subspace of martingales.

The subspace of martingales must be contained in the range of an operator that maps from L values to the space of edge flows. Denote this operator $M(\beta)$ since it maps to the space of martingales. In detailed balance $M = RC^{\intercal}$. Our goal is to find a first order perturbation to this operator that defines the space of martingales in the weak rotation limit.

Expand $M(\beta) = R\left[C^{\intercal} + \beta B^{\intercal}\right] + \mathcal{O}(\beta^2)$. Then, the range of $M(\beta)$ is automatically orthogonal to the range of possible fluxes in the limit β vanishes if the matrix B satisfies:

$$BG = CF_{\text{rot}}H$$

or, equivalently:

$$G^{\intercal}B^{\intercal} = H^{\intercal}F_{\text{rot}}C^{\intercal}. \tag{7.77}$$

The matrix B is $L \times E$ and the matrix $G^{\intercal}$ is $V \times E$. Therefore this is set of VL equations in EL unknowns. Since $E > V$ for all cases of interest, B is not uniquely specified by Equation (7.77). Equation (7.77) does not uniquely define B since there are always infinitely many matrices with the same range, so we should only expect B to be restricted to a subspace of possible matrices.

A natural way to solve for a unique B is to restrict B to be minimal in some norm. For example, suppose that we look for a matrix B that satisies the orthogonality constraint exactly, while having the smallest possible Frobenius norm. then each row of B is the solution to a constrained optimization problem. Let b_l denote the l^{th} row of B. Then, to minimize the Frobenius norm of B, b_l must minimize $||b||_2$ subject to the constraint

$G^\intercal b = [H^\intercal F_{\text{rot}} C^\intercal]_l$ where $[H^\intercal F_{\text{rot}} C^\intercal]_l$ denotes the l^{th} column of $H^\intercal F_{\text{rot}} C^\intercal$. The divergence maps from the space of edges to the space of nodes. The space of edges is either larger in dimension than the space of nodes (provided the graph is not a tree), or equal in dimension to the space of nodes. This means that the constraint equation has more unknowns than equations, so will typically be satisfied anywhere inside an affine subspace. The subspace satisfying the constraint can be written $b_* \oplus \text{null}\{G^\intercal\}$ where b_* is a particular solution to the constraint equation, provided b_* exists. To show that b_* exists we must show that there is at least one solution to the orthogonalist constraint equation.

The orthogonality constraint has a solution if $[H^\intercal F_{\text{rot}} C^\intercal]_l$ is in the range of the divergence, $G^\intercal$ for all l. This requires that it does not have a projection onto the null space of G for any l. The null-space of G is the range of the vector of all ones $\mathbf{1}$, so, in order for the orthogonality constraint to be satisfied the inner product $\mathbf{1}^T [H^\intercal F_{\text{rot}} C^\intercal]_l = 0$. That is, the sum of the entries of $[H^\intercal F_{\text{rot}} C^\intercal]_l$ must equal zero for all l. That is, all the columns of $H^\intercal F_{\text{rot}} C^\intercal$ must sum to zero. For now we assume this is possible, and show how to solve for B accordingly. In fact this is not possible, since the sum of the columns of $H^\intercal F_{\text{rot}} C^\intercal$ is the same as the sum of the rows of $C F_{\text{rot}} H$ which equals $C F_{\text{rot}} H \mathbf{1} = 2C f_{\text{rot}} \neq \mathbf{0}$. We will show that this is the extra direction which is in the space of martingales in detailed balance, but must be removed to find the space of martingales near detailed balance.

Suppose it were possible to solve for B. If B solves the constrained minimization problem then the rows of B must be in the range of G. Any solution to a constrained minimization problem occurs at points where the gradient of the cost function is zero when projected back onto the subspace defined by the constraint. This implies that the gradient of the cost function must be orthogonal to the subspace defined by the constraint at the minimizer. Since our cost function is $||b||_2^2$, the gradient of the cost function is $2b$. Given any b that satisfies the constraint, the constraint subspace can always be expressed as $b +$

null$\{G^\intercal\}$. This implies that the solution to the minimization problem occurs at the point b such that $G^\intercal b$ satisfies the constraint, and b is orthogonal to the nullspace of the divergence, $G^\intercal$. The nullspace of the divergence is the range of the curl transpose, therefore the vector space orthogonal to the nullspace of the divergence is the range of the gradient.

It follows that the matrix B with minimal Frobenius norm that defines the space of martingales in the weak rotation limit has rows:

$$\begin{aligned} b_l &= -G\phi_B(l) \\ G^\intercal G\phi_B(l) &= -[H^\intercal F_{\text{rot}} C^\intercal]_l. \end{aligned} \tag{7.78}$$

where $\Phi_B \in \mathbb{R}^{L\times V}$ and $B^\intercal = -G\Phi_B$.

Equation (7.78) implies that, for B with minimal Frobenius norm, the perturbation $\beta R^{-1}B$ is orthogonal with respect to R^{-2} to $M(0)$:

$$\lim_{\epsilon\to 0} \frac{1}{\beta} M(0)^\intercal R^{-2}(M(\beta) - M(0)) = -CRR^{-2}RG\Phi_B = -CG\Phi_B = 0.$$

Therefore the choice of minimization in the Frobenius norm is natural since it ensures that the perturbation B contains entirely new information (in the metric R^{-2}) about the vector space of martingales in weak rotation relative to the vector space of martingales in detailed balance.

Notice that Equation (7.78) transforms the problem of finding B into a sequence of L discrete Poisson equations. So, despite venturing far afield from the original setting of the HHD we recover, yet again, a concrete link between dynamics of a process near detailed balance and HHD type equations.

To make sense of the right hand side of the discrete Poisson equation, Equation (7.78), consider the i, l entry of $H^\intercal F_{\text{rot}} C^\intercal$. Here i indexes a particular node and l indexes a partic-

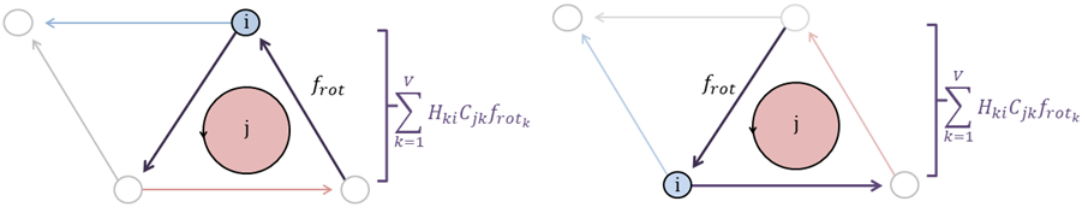

Figure 7.4: Schematic showing the ij entry of $CF_{\text{rot}}H$. The product $CF_{\text{rot}}H$ forms the right hand side of the discrete Poisson equation for the perturbation to the subspace of martingales. Here i is shown for two possible nodes, and j for a single loop. The edges where $H_{ki} \neq 0$ are shown in blue, and the edges where $C_{jk} \neq 0$ are shown in red. The edges where these two sets overlap are shown in purple. The right hand side is given by summing f_{rot} over these edges.

ular basis loop. The corresponding entry is given by a sum over the edges. Specifically:

$$[H^\intercal \text{diag}(f_{\text{rot}})C^\intercal]_{il} = \sum_{k=1}^{V} H_{ki}C_{lk}f_{\text{rot}k}.$$

The entries of H are all either one or zero, and $h_{ki} = 1$ if and only if the k^{th} edge is adjacent to node i. Therefore, multiplication by $H^\intercal$ is equivalent to restricting the sum over edges to the edge neighborhood of i. Denote the set of all edges neighboring i: $\mathcal{E}(i)$. Then:

$$[H^\intercal \text{diag}(f_{\text{rot}})C^\intercal]_{il} = \sum_{k \in \mathcal{E}(i)} C_{lk}f_{\text{rot}k}.$$

Thus the right hand side of the l^{th} discrete Poisson equation is a vector with an entry for every node, with that entry set to the path integral of f_{rot} along the segment of the l^{th} basis loop passing through node i. This is large at a particular i, l if the flow of f_{rot} around the l^{th} loop is large in the neighborhood of node i.

So, given a system near detailed balance, if we could solve for B it would be possible to find a perturbation to the operator $M(\beta)$ that maps to a subspace containing the space of martingales in small β by solving a sequence of discrete Poisson equations with right

hand side that depends on the rotational component of the edge flow, before scaling by the resistances. However, as noted at the start of this analysis, the subspace spanned by $M(\beta)$ has an extra dimension, so will typically include edge flows which are not martingales in its range.

Suppose that the probability distribution p is set to the stationary distribution $q(\beta)$. Then, to first order, the steady state fluxes are:

$$J(\beta) = \beta R^{-1} \left(G\hat{q}^{(1)} + 2f_{\text{rot}} \right) + \mathcal{O}(\beta^2).$$

Then, to first order:

$$M(\beta)^{\intercal} J(\beta) = \beta[C + \beta B] R R^{-1} [G\hat{q}^{(1)} + 2f_{\text{rot}}] = \beta[CG\hat{q}^{(1)} + 2Cf_{\text{rot}}] + \mathcal{O}(\beta^2).$$

The product CG vanishes since the curl and gradient are orthogonal, leaving:

$$M(\beta)^{\intercal} J(\beta) = 2\beta C f_{\text{rot}} + \mathcal{O}(\beta^2). \tag{7.79}$$

The curl of the rotational field is never zero if the process is outside of detailed balance, otherwise the rotational potential associated with the HHD would not be uniquely defined. Therefore, regardless of our careful choice of the first order perturbation B, there exists a current, the steady state current $J(\beta)$, such that the product $M(\beta)^{\intercal} J(\beta)$ is still first order in β. This implies that there exists a edge flow in the subspace spanned by $M(\beta)$ that does not correspond to a martingale to first order in β. This contradicts the claim that the range of $M(\beta)$ corresponds to the subspace that is orthogonal to the range of currents up to first order in β. Perhaps even more surprisingly, this contradiction holds for any choice of B, since B does not appear in the first order terms of the expansion shown above.

To see what went wrong it is helpful to write the expansion in more concise notation. Let $M(\beta) = M^{(0)} + M^{(1)}\beta + \mathcal{O}(\beta^2)$ and let the operator mapping from p to J (probability to fluxes) be denoted $N(\beta) = N^{(0)} + \beta N^{(1)} + \mathcal{O}(\beta^2)$. Then, applied to a distribution p:

$$M(\beta)^\intercal N(\beta)p = M^{(0)\intercal}N^{(0)}p + \beta(M^{(0)\intercal}N^{(1)} + M^{(1)\intercal}N^{(0)})p + \mathcal{O}(\beta^2).$$

Setting each order to zero for all p gives the usual equations $M^{(0)\intercal}N^{(0)} = 0$ and $M^{(1)\intercal}N^{(0)} = -M^{(1)\intercal}N^{(1)}$. The previous analysis solved for $M^{(0)}$ and $M^{(1)}$ satisfying these constraints given the particular forms of $N^{(0)}$ and $N^{(1)}$ that arise in the weak rotation limit.

Replace p with the steady state $q(\beta)$. Then:

$$M(\beta)^\intercal N(\beta)q(\beta) = M^{(0)\intercal}N^{(0)}q^{(0)} + \beta\left[(M^{(0)\intercal}N^{(1)} + M^{(1)\intercal}N^{(0)})q^{(0)} + M^{(0)\intercal}N^{(0)}q^{(1)})\right] + \mathcal{O}(\beta^2).$$

By definition $q^{(0)}$ is the equilibrium distribution of the system in the limit that β goes to zero. Therefore $N^{(0)}q^{(0)} = 0$. We enforced $M^{(0)\intercal}N^{(0)} = 0$ to find the space of martingales in detailed balance. This leaves:

$$M(\beta)^\intercal N(\beta)q(\beta) = \beta M^{(0)\intercal}N^{(1)}q^{(0)} + \mathcal{O}(\beta^2).$$

It is easy to check that this remainder is precisely the term that survived when taking the inner product with the steady state currents. The remainder is independent of $M^{(1)}$, which was determined by the perturbation B, so is independent of the choice of the perturbation to the matrix that maps to the space of martingales. Therefore the product is $\mathcal{O}(\beta)$ unless $M^{(0)\intercal}N^{(1)}q^{(0)} = 0$.

This introduces an additional constraint on $M^{(0)}$ that we had not enforced before. Originally we solved for $M^{(0)}$ such that $M^{(0)\intercal}N^{(0)} = 0$. Since $N^{(0)\intercal}$ had a nullspace with dimension L it was possible to find a matrix $M^{(0)}$ with rank L, namely $C^\intercal$, such that $M^{(0)\intercal}N^{(0)} = 0$. We then solved for a first order perturbation to this matrix, $M^{(1)}$, by enforcing $M^{(1)\intercal}N^{(0)} = M^{(0)\intercal}N^{(1)}$. However, since the dimension of the space of martingales should have been $L-1$ as soon as $\beta \neq 0$, the range of $M^{(0)}+\beta M^{(1)}$ included an extra dimension. If, however, we had enforced the additional constraint, $M^{(0)\intercal}N^{(1)}q^{(0)} = 0$, then, provided the constraint is unique from the constraints $M^{(0)\intercal}N^{(0)} = 0$, then any mapping satisfying the contraints must have range $L-1$ instead of L.

To remove the extra dimension set:

$$M^{(0)} = RC^\intercal Z \tag{7.80}$$

where Z is $L \times L-1$, has orthonormal columns, and satisfies:

$$Z^\intercal C f_{\text{rot}} = 0. \tag{7.81}$$

Then Z must be an orthonormal basis for the set of functions on the loops that are orthogonal to Cf_{rot}. Then, if we let B be a $E \times L-1$ matrix, the new orthogonality constraint reads:

$$G^T B^T = H^T F_{\text{rot}} C^T Z \tag{7.82}$$

which has a solution provided the columns of $H^T F_{\text{rot}} C^T Z$ sum to zero. The columns sum to zero if the rows of $Z^T C F_{\text{rot}} H$ sum to zero, which requires $Z^T C f_{\text{rot}} = 0$, which is necessarily true by the construction of Z. Thus, by restricting $M^{(0)}$ with Z there is a necessarily a solution to the orthogonality constraint on each row of B. Then each row is

the solution to the Poisson equation:

$$\begin{aligned} b_l &= -G\phi_B(l) \\ G^\intercal G\phi_B(l) &= -[H^\intercal F_{\text{rot}} C^\intercal Z]_l. \end{aligned} \tag{7.83}$$

Therefore, the space of martingales is the range of $M(\beta)$, which, to first order in β, equals:

$$M(\beta) = R^{-1}\left[C^\intercal Z + \beta B^\intercal\right] + \mathcal{O}(\beta^2). \tag{7.84}$$

The product with Z ensures that the operator $M(\beta)$ is $E \times L - 1$, and has rank $L - 1$ rather than L. The introduction of Z to the operator spanning the space of martingales is easy to understand qualitatively. Suppose we had not added Z. Then we could have defined an observable with field s set to $C^\intercal\theta$. Then the zeroeth order term in the observable would have been $R^{-1}f_{\text{rot}}$ so the steady state production of the observable would have been $||f_{\text{rot}}||_2^2$. This production occurs because the steady state currents tend to move in the direction of f_{rot}. Therefore, if the field also aligned with f_{rot} the observable would tend to increase with time. To ensure that the observable is a martingale the field cannot align with this small circulating current, therefore Z needs to be introduced to restrict the range of $C^\intercal$ to rotational fields orthogonal to f_{rot}.

In summary, the space of martingales is the range of the mapping $M(\beta)$, which up to first order in β, can be computed by:

1. Solve for the scalar potential, associated equilibrium and R
2. Solve the f_{rot} and for a basis $Z \in \mathbb{R}^{L\times L-1}$ such that $Z^\intercal[CC^\intercal]\theta = 0$.
3. One row at a time solve the discrete Poisson equation for B (Equation (7.78))
4. Then set $M(\beta) = R\left[C^\intercal Z + \beta B^\intercal\right] + \mathcal{O}(\beta^2)$.

The Weak Rotation Expansion: Convergence and the Weak Rotation Regime

The full weak rotation expansion is defined by the recursive sequence of equations:

$$\begin{aligned}
\hat{L}^{(0)}\hat{q}^{(0)} &= 0 \\
\hat{L}^{(0)}\hat{q}^{(1)} &= -\hat{L}^{(1)}\hat{q}^{(0)} \\
&\vdots \\
\hat{L}^{(0)}\hat{q}^{(n)} &= -\sum_{j=1}^{n} \hat{L}^{(j)}\hat{q}^{(n-j)}
\end{aligned}$$

where:

$$\hat{L}^{(n)} = -G^{\intercal}R^{-1}F_{\text{rot}}^{n}K^{(n)} \text{ given } K^{(n)} = \begin{Bmatrix} G \text{ if } n \text{ even} \\ H \text{ if } n \text{ odd} \end{Bmatrix}$$

and where $\hat{q}(\beta)$ is the steady state of the nonequilibrium process scaled by the equilibrium distribution of the corresponding conservative process. Then, at any order n, the corresponding correction to the steady state obeys a weighted Poisson equation of the form:

$$G^{\intercal}R^{-1}G\hat{q}^{(n)} = -G^{\intercal}\sum_{j=1}^{n}\frac{1}{j!}R^{-1}F_{\text{rot}}^{j}K^{(j)}\hat{q}^{(n-j)} \tag{7.85}$$

where the right hand side can be interpretted as the divergence of an edge flow defined by a sum over the lower order corrections. Removing the divergence from the Laplacian produces the mapping from probabilities to fluxes so the steady state fluxes are:

$$J^{(n)} = R^{-1}\left[G\hat{q}^{(n)} + \sum_{j=1}^{n}\frac{1}{j!}F_{\text{rot}}^{j}K^{(j)}\hat{q}^{(n-j)}\right] \tag{7.86}$$

which is the difference between the edge flow on the right hand side of Equation (7.85) and its approximation with the gradient of $\hat{q}^{(n)}$. Therefore the n^{th} order corrections to the

steady state and fluxes satisfy the same weighted HHD as the first order corrections, only with a different right hand side that is defined recursively:

$$-G\hat{q}^{(n)} + RC^{\intercal}\theta_J^{(n)} = -G\hat{q}^{(n)} + RJ^{(n)} = \sum_{j=1}^{n} \frac{1}{n!} F_{\text{rot}}^{j} K^{(j)} \hat{q}^{(n-j)}. \tag{7.87}$$

Equation (7.87) shows that the Taylor expansion of the steady state and the steady state fluxes is a recursive sequence of weighted Helmholtz-Hodge decompositions. The left hand side is the same at all orders, so the bulk of the theory developed for the first order corrections extends to all order corrections with f_{rot} replaced by $\sum_{j=1}^{n} \frac{1}{n!} F_{\text{rot}}^{j} K^{(j)} \hat{q}^{(n-j)}$. As in the first order case, at the n^{th} order the scalar part of the HHD is associated with the correction to the stationary distribution and the rotating part is associated with the correction to the steady state fluxes. At every order there is a trade-off between driving circulation and perturbing the steady state. The more efficiently current is driven the less the steady state changes at all orders. Thus the fundamental exchange between producing current and perturbing the steady state is true at all orders, not just first order.

A key difference at higher orders is that f_{rot}^{n} is rarely divergence free for $n > 1$ even though f_{rot} is divergence free. Suppose two loops overlap in the same direction. Then, on the shared edge $f_{\text{rot}}^{n} = (\theta_I + \theta_{II})^n$ will grow faster than the sum of $\theta_I^n + \theta_{II}^n$, often producing a bottleneck at the lee end of the shared boundary. Therefore there may be bottlenecks where the divergence of the right hand side is nonzero even if all the resistances are the same. Note that this mirrors the difficulty identified in Section 7.2.3 that makes finding the nonequilibrium steady state hard to solve.

It is important to note that the weighted HHD equation/weighted Poisson equations only define $\hat{q}^{(n)}$ up to a constant. In order to solve for a unique $\hat{q}^{(n)}$ we require that $\hat{q}(\beta)$ is normalized for all β. If $\hat{q}(\beta)$ is normalized for all β then $\frac{d^n}{d\beta^n} \sum_{i=1}^{V} \hat{q}_i(\beta) = 0$ for any β. In

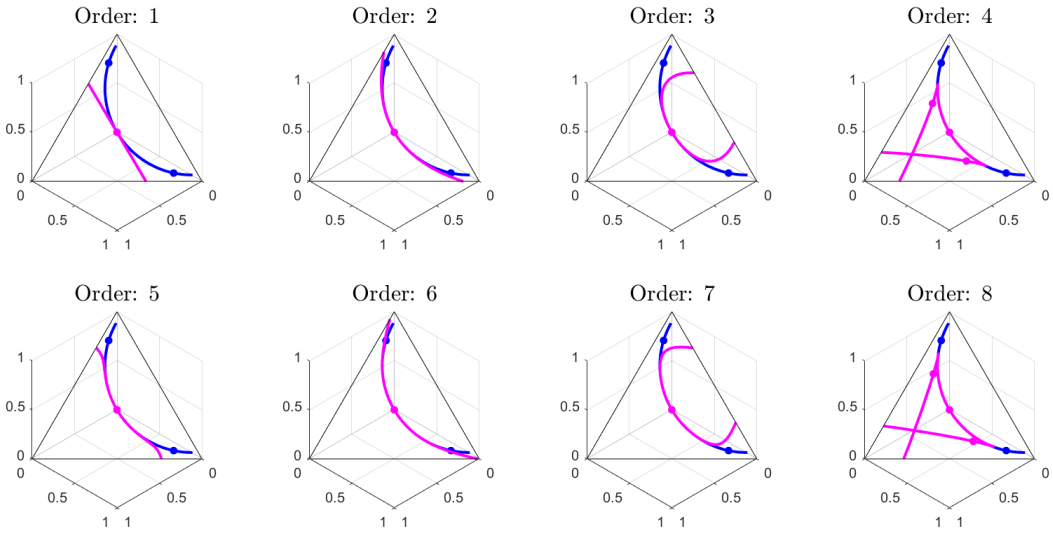

Figure 7.5: The first eight approximations to the steady state of the scaled (purely rotational) system as a function of β for a nonequilibrium process on a triangular network with $\theta = 1$. The exact steady states are shown in blue, and the approximations are shown in magenta. The central magenta dot represents the steady state at $\beta = 0$. The blue outer dots represent the exact solutions for $\beta = \pm 1$. The corresponding magenta dots (which are not visible at all orders with this axis scaling) are the approximations at $\beta = \pm 1$.

particular, if $\beta = 0$, then the n^{th} derivative is the n^{th} term in the expansion so $\sum_{i=1}^{V} \hat{q}_i^{(0)} = 1$ and $\sum_{i=1}^{V} \hat{q}_i^{(n)} = 0$ for all $n > 0$. Therefore, for $n > 0$ the correction $\hat{q}^{(n)}$ is chosen so that it solves the weighted Poisson equation and so that the sum of $\hat{q}^{(n)}$ equals zero. This can be accomplished by first solving for a particular solution to the weighted Poisson equation, computing the sum of the solution, and subtracting it off.

Performing the recursive sequence Equation (7.87) out n steps gives the n^{th} order Taylor expansion of the steady state $\hat{q}(\epsilon)$, and the steady state currents J.

A series of approximations of this type are shown in Figure 7.5 for a three state loop with scalar potential $\phi = [-1, 2, 3]/4$ and conductances $\rho = [1, 0.1, 1]$. The exact steady states are shown in blue, and the approximations are shown in magenta. The central magenta dot represents the steady state at $\beta = 0$. The blue outer dots represent the exact

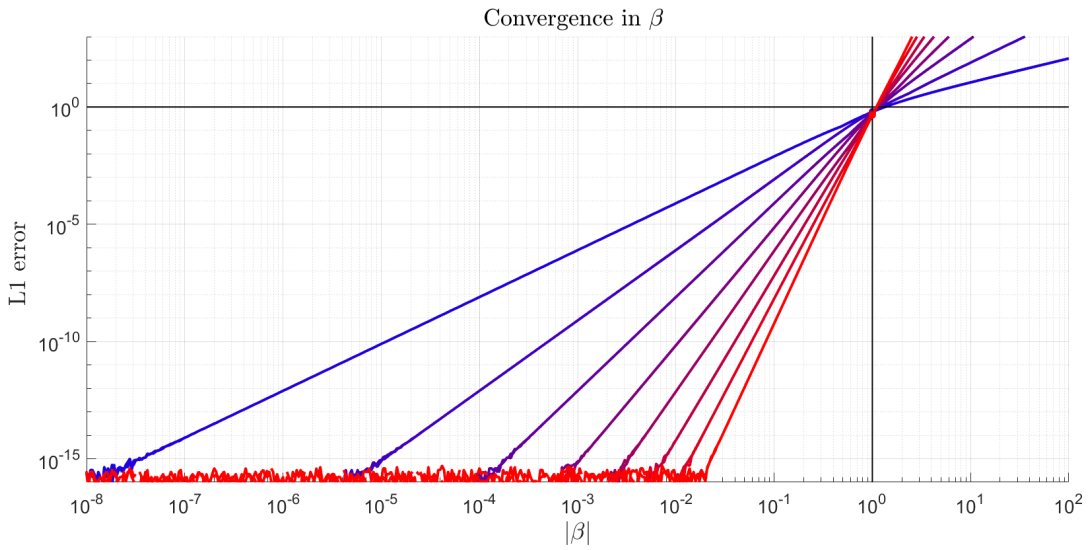

Figure 7.6: Convergence of the first eight approximations to the steady state of the approximation shown in Figure 7.5. The blue curve represents the first order approximation and the red curve represents the eighth order expansion. The slope of each matches the expected rate of convergence. For example, the eighth order expansion converges $\mathcal{O}(\beta^9)$ so has slope 9 on the convergence plot.

solutions for $\beta = \pm 1$. The corresponding magenta dots are the approximations at $\beta = \pm 1$. Notice that the approximation at $\beta = \pm 1$ cycles periodically about the exact steady state. The convergence of the approximation at each order are shown in Figure 7.6.

The convergence experiment was performed out to order 25, and for every order the error in the approximation decayed proportional to β^{n+1} for $\beta < 1$. It is unreasonable to expect the expansion to converge for all β since the distribution always has to be normalized. The expansion expresses the steady state at each vertex as a power series in β. Since the distribution must be normalized and nonnegative at all vertices the power series must remain bounded for $\beta \to -\infty$ and $\beta \to \infty$.

In order to show that the recursive expansion is valid we need to show that it converges to $q(\beta)$ for some $\beta > 0$. Since the expansion is a power series it suffices to show that the radius of convergence of the power series is greater than zero. An essential goal is to find

a nonzero lower bound on the radius of convergence of this expansion. The weak rotation regime is the space of f_{rot} such that the weak rotation expansion converges.

Theorem 32 (The Weak Rotation Expansion)**.** *If $L(\beta)$ is an arbitrary Laplacian for a Markov process obeying microscopic reversibility on a finite network where $l_{ji}(\beta)$ equals $\rho_{ij}\exp\left(f_{conij}+\beta f_{rotij}\right)$ then, after scaling to a purely rotational process via the transform defined in Theorem 25, the n^{th} order terms in the Taylor expansion of the steady state and steady state current are solutions to the recursive sequence of HHD's:*

$$\hat{q}^{(0)}=\boldsymbol{1},\quad -G\hat{q}^{(n)}+RC^{\intercal}\theta_J^{(n)}=-G\hat{q}^{(n)}+RJ^{(n)}=\sum_{j=1}^{n}\frac{1}{n!}F_{rot}^{j}K^{(j)}\hat{q}^{(n-j)}.$$

The expansion has a finite radius of convergence, and converges if $\beta||f_{rot}||_1\leq 2/3$ or $\beta||f_{rot}||_\infty\leq\frac{1}{Vd+1/2}$ where d is the maximum degree of any vertex in the network.

Proof. The expansion is of the form:

$$\hat{q}(\beta)=\sum_{k=0}^{\infty}\hat{q}^{(k)}\beta^k$$

which converges if:

$$||\hat{q}(\beta)||_\infty\leq\sum_{k=0}^{\infty}||\hat{q}^{(k)}||_\infty\beta^k<\infty. \tag{7.88}$$

Therefore we will attempt to derive an upper bound $||\hat{q}^{(k)}||_\infty\leq a_k$. Convergence in the infinity norm implies convergence at all vertices j. The power series $\sum_{k=0}^{\infty}\hat{q}_j^{(k)}\beta^k$ converges if it converges absolutely. It converges absolutely if $\sum_{k=0}^{\infty}|\hat{q}_j^{(k)}|\beta^k$ converges. By definition of the infinity norm $|\hat{q}_j^{(k)}|<\max_j\{|\hat{q}_j^{(k)}|\}=||\hat{q}^{(k)}||_\infty$ so:

$$\sum_{k=0}^{\infty}|\hat{q}_j^{(k)}|\beta^k\leq\sum_{k=0}^{\infty}||\hat{q}^{(k)}||_\infty\beta^k.$$

Therefore convergence of the power series with coefficients $||\hat{q}^{(k)}||_\infty$ implies convergence of the weak rotation expansion at every node. This can be generalized to the other induced matrix norms by noting that $||\hat{q}^{(k)}||_\infty$ is less than or equal to $||\hat{q}^{(k)}||_1$ and $||\hat{q}^{(k)}||_2$. Thus convergence of the power series with coefficients $||\hat{q}^{(k)}||_{1,2,\infty}$ implies that convergence of the weak rotation expansion at every node.

The k^{th} order term in the expansion satisfies the weighted HHD:

$$G\hat{q}^{(k)} + RC^\intercal\hat{\theta}^{(k)} = \sum_{j=1}^{k} \frac{1}{k!} F_{\text{rot}}^k K^{(k)} \hat{q}^{(k-j)} = f^{(k)}.$$

With the constraint $\sum_i \hat{q}_i^{(k)} = 0$ if $k > 0$. Therefore $||\hat{q}^{(k)}||_\infty \leq \max_{i,j} |\hat{q}_i^{(k)} - \hat{q}_j^{(k)}|$ where the maximum runs over all pairs of nodes in the network (this includes pairs that are not connected by one edge). Since the correction $\hat{q}^{(k)}$ satisfies a weighted HHD the difference in $\hat{q}_i^{(k)}$ and $\hat{q}_j^{(k)}$ is the average work over all paths from i to j, sampled according to a simple random walk with weights R^{-1} (see Theorem 12). As usual the set of paths can be restricted to paths without loops since there always exists a set of $L+1$ simple paths (no loops and no backtracking) such that a weighted average of the work over these paths gives the difference in $\hat{q}_i^{(k)}$ and $\hat{q}_j^{(k)}$. Then, since the work is a sum of $\pm f^{(k)}$ over paths of length less than or equal to E:

$$||\hat{q}^{(k)}||_\infty \leq ||f^{(k)}||_1. \tag{7.89}$$

To bound $||f^{(k)}||_1$ use the triangle inequality and definition of the induced matrix norm:

$$||f^{(k)}||_1 \leq \sum_{j=1}^{k} \frac{1}{j!} ||F_{\text{rot}}^j||_1 ||K^{(j)}||_1 ||\hat{q}^{(k-j)}||_1.$$

To simplify, note that $||\hat{q}^{(k-j)}||_1 \leq V||\hat{q}^{(k-j)}||_\infty$. The matrix F_{rot}^k is diagonal, so $||F_{\text{rot}}^k||_1$ is $||f_{\text{rot}}||_\infty^k$. The one norm of a matrix is equal to its maximum absolute column sum. The

absolute value of G (entrywise) is H so $||K^{(k)}||_1 = ||H||_1$. The columns of H correspond to particular nodes, and the ij entries are equal to one if i and j are neighbors. Therefore $||H||_1 = d$ where d is the maximum degree of all the nodes in the network.

Therefore:

$$||\hat{q}^{(k)}||_\infty \leq \sum_{j=1}^{k} \frac{Vd}{k!} ||f_{\text{rot}}||_\infty^k ||\hat{q}^{(k-j)}||_\infty. \tag{7.90}$$

Define the recursive sequence:

$$a_k = \sum_{j=1}^{k} \frac{Vd}{j!} ||f_{\text{rot}}||_\infty^j a_{k-j}, \quad a_0 = 1/V.$$

Then:

$$||\hat{q}^{(k)}||_\infty \leq a_k.$$

To simplify, define the coefficients:

$$b_k(Vd) \frac{||f_{\text{rot}}||_\infty^k}{V} = a_k.$$

Then:

$$b_k(Vd) = \sum_{j=1}^{k} \frac{Vd}{j!} b_{k-j}(Vd), \quad b_0(Vd) = 1.$$

Since:

$$\begin{aligned} b_k(Vd) = & \frac{V}{||f_{\text{rot}}||_\infty^k} a_k = \frac{V}{||f_{\text{rot}}||_\infty^k} \sum_{j=1}^{k} \frac{Vd}{j!} ||f_{\text{rot}}||_\infty^j a_{k-j} = \\ & \frac{V}{||f_{\text{rot}}||_\infty^k} \sum_{j=1}^{k} \frac{Vd}{j!} ||f_{\text{rot}}||_\infty^j b_{k-j}(Vd) \frac{||f_{\text{rot}}||_\infty^{k-j}}{V} = \\ & \frac{V}{||f_{\text{rot}}||_\infty^k} \frac{||f_{\text{rot}}||_\infty^k}{V} \sum_{j=1}^{k} \frac{Vd}{j!} b_{k-j}(Vd) = \sum_{j=1}^{k} \frac{Vd}{j!} b_{k-j}(Vd). \end{aligned}$$

Let $x = Vd$. Notice that $x > 1$ since $V > 1$ and $d > 1$ for any network with a loop.

Then the collection of coefficients $b_k(x)$ is a recursive sequence of polynomials in x defined by the recursion:

$$b_k(x) = x \sum_{j=1}^{k} \frac{b_{k-j}(x)}{j!}, \quad b_0(x) = 1. \tag{7.91}$$

Notice that $b_k(x)$ is a monic polynomial of order k in x. The terms of the polynomial bound the perturbation in the steady state:

$$||\hat{q}^{(k)}||_\infty \leq \frac{b_k(x)}{V} ||f_{\text{rot}}||_\infty^k. \tag{7.92}$$

Therefore, the n^{th} order expansion of the steady state can be bounded above by:

$$||\sum_{k=0}^{n} \hat{q}^{(k)} \beta^k||_\infty \leq \sum_{k=0}^{n} ||\hat{q}^{(k)}||_\infty \beta^k \leq \sum_{k=0}^{n} \frac{b_k(x)}{V} ||f_{\text{rot}}||_\infty^k \beta^k. \tag{7.93}$$

The expansion converges absolutely if the power series $\sum_{k=0}^{n} \frac{b_k(x)}{V} ||f_{\text{rot}}||_\infty^k \beta^k$ converges as n goes to infinity. Therefore the radius of convergence for the expansion of the steady state is greater than or equal to the radius of convergence of the power series defined by $\sum_{k=0}^{n} \frac{b_k(x)}{V} ||f_{\text{rot}}||_\infty^k \beta^k$. To find the radius of convergence of the power series use the ratio test. This requires taking the limit:

$$\lim_{k\to\infty} \frac{\frac{b_{k+1}(x)}{V} ||f_{\text{rot}}||_\infty^{k+1} |\beta|^{k+1}}{\frac{b_k(x)}{V} ||f_{\text{rot}}||_\infty^k |\beta|^k} = \lim_{k\to\infty} \frac{b_{k+1}(x)}{b_k(x)} ||f_{\text{rot}}||_\infty \beta.$$

A lower bound on the radius of convergence requires an upper bound on the limit of the ratio of $b_{k+1}(x)$ and $b_k(x)$. From the recursion:

$$\frac{b_{k+1}(x)}{b_k(x)} = x \frac{\sum_{j=1}^{k+1} b_{k+1-j}(x)/j!}{b_k(x)} = x \sum_{j=0}^{k} \frac{b_{k-j}(x)}{(j+1)!\, b_k(x)} = x \left[\frac{b_k(x)}{b_k(x)} + \sum_{j=1}^{k} \frac{b_{k-j}(x)}{(j+1)!\, b_k(x)} \right].$$

Substitute in the recursion for $b_k(x)$:

$$\frac{b_{k+1}(x)}{b_k(x)} = x\left[1 + \frac{1}{x}\frac{\sum_{j=1}^{k}\frac{1}{(j+1)!}b_{k-j}(x)}{\sum_{j=1}^{k}\frac{1}{j!}b_{k-j}(x)}\right].$$

Then:

$$\lim_{k\to\infty}\frac{b_{k+1}(x)}{b_k(x)} = x + \lim_{k\to\infty}\frac{\sum_{j=1}^{k}\frac{1}{(j+1)!}b_{k-j}(x)}{\sum_{j=1}^{k}\frac{1}{j!}b_{k-j}(x)}. \tag{7.94}$$

This is easily bounded by noting that $\frac{1}{(j+1)!}b_{k-j}(x) < \frac{1}{(j)!}b_{k-j}(x)$ so the ratio is strictly less than one. More strongly, since j starts from one, $\frac{1}{(j+1)!}b_{k-j}(x) < \frac{1}{2}\frac{1}{(j)!}b_{k-j}(x)$ Therefore:

$$\lim_{k\to\infty}\frac{b_{k+1}(x)}{b_k(x)} \leq x + \frac{1}{2}. \tag{7.95}$$

It follows that the limit of the original ratio is (plugging in $x = Vd$):

$$\lim_{k\to\infty}\frac{\frac{b_{k+1}(x)}{V}||f_{\text{rot}}||_\infty^{k+1}|\beta|^{k+1}}{\frac{b_k(x)}{V}||f_{\text{rot}}||_\infty^{k}|\beta|^{k}} \leq \left(Vd + \frac{1}{2}\right)||f_{\text{rot}}||_\infty\beta.$$

To guarantee convergence this ratio must be less than one, so:

$$\beta < \frac{1}{(Vd + \frac{1}{2})||f_{\text{rot}}||_\infty}. \tag{7.96}$$

For a given Laplacian $\beta = 1$ so Equation (7.96) requires:

$$||f_{\text{rot}}||_\infty < \frac{1}{Vd + \frac{1}{2}}. \tag{7.97}$$

Then Equation (7.97) implies the radius of convergence of the weak rotation expansion,

R, is bounded below by:

$$R > \frac{1}{Vd + \frac{1}{2}} > 0. \tag{7.98}$$

so the weak rotation expansion has a nonzero radius of convergence, and is absolutely convergent to $\hat{q}(\beta)$ for $\beta||f_{\text{rot}}||_\infty$ sufficiently small.

Unfortunately this lower bound on the radius of convergence is likely an extreme underestimate when V is big. For example, in the numerical test presented earlier $V = 3$ and $d = 2$ so the radius of convergence had to be greater than $1/7$, but we observed convergence for $\beta \leq 1$. Thus the radius of convergence is likely much larger than the lower bound given by Equation (7.98). The factor of V entered the analysis when bounding $||\hat{q}^{(k-j)}||_1$ above by $||\hat{q}^{(k-j)}||_\infty$. It is possible that it could be eliminated by starting with $||\hat{q}(\beta)||_1$ or $||\hat{q}(\beta)||_2$.

Consider $||f_{\text{rot}}||_1$ instead of $||f_{\text{rot}}||_\infty$. Using $||f_{\text{rot}}||_1$ avoids the direct dependence on V, and leads to a more elegant conclusion. That said, it still requires that $||f_{\text{rot}}||_\infty$ be small relative to $1/E$, so whether or not this is a tighter bound will depend on the distribution of degrees of the nodes.

In the previous analysis we introduced a bound for the one-norm of $F^k_{\text{rot}} K^{(k)} \hat{q}^{(k-j)}$. We used the induced one norm for matrices to derive the bound $||F^k_{\text{rot}} K^{(k)} \hat{q}^{(k-j)}||_1$ is less than or equal to $||f^k_{\text{rot}}||_\infty d||\hat{q}^{(k-j)}||_1$. The factor of V was introduced to change back into the infinity norm. This factor of V can be avoided by noting that $||F_{\text{rot}} K^{(k)} \hat{q}^{(k-j)}||_1$ is a sum over the edges of f^k_{rot} on the edge times either the sum or the difference of $\hat{q}^{(k-j)}$ at the endpoints of the edge. This sum or difference is always strictly less than $2||\hat{q}^{(k-j)}||_\infty$. Therefore:

$$||F_{\text{rot}} K^{(k)} \hat{q}^{(k-j)}||_1 \leq 2||f^k_{\text{rot}}||_1 ||\hat{q}^{(k-j)}||_\infty. \tag{7.99}$$

To finish:

$$||f^k_{\text{rot}}||_1 = \sum_{ij} |f^k_{\text{rot}ij}| \leq \left(\sum_{ij} |f_{\text{rot}ij}| \right)^k = ||f_{\text{rot}}||^k_1. \tag{7.100}$$

Now:

$$||\hat{q}^{(k)}||_\infty \leq \sum_{j=1}^{k} \frac{1}{k!} ||f_{\text{rot}}||^k_1 ||\hat{q}^{(k-j)}||_\infty. \tag{7.101}$$

This leads to the same recursion as before, only $x = 1$ instead of Vd and $||f_{\text{rot}}||_\infty$ is replaced with $||f_{\text{rot}}||_1$. Therefore the weak rotation expansion converges provided:

$$\beta ||f_{\text{rot}}||_1 \leq \frac{1}{1 + 1/2} = \frac{2}{3}. \tag{7.102}$$

□

At face value the bound established using the one norm appears tighter since it does not appear to depend on V or d. However, to compare to the original bound in terms of the infinity norm we need to set an upper limit on $||f_{\text{rot}}||_\infty$ not $||f_{\text{rot}}||_1$. This is typically preferable since it means that the weak rotation expansion converges provided the rotational force at each edge is small, not that the sum of the rotational forces over all the edges is small. This sum will depend on the number of edges, so for a fixed average rotational force, but growing network, a bound on $||f_{\text{rot}}||_1$ may not be enough to prove convergence. Therefore, in terms of the average rotational force $\bar{f}_{\text{rot}} = ||f_{\text{rot}}||_1 / E$ Equation (7.102) requires:

$$\beta \bar{f}_{\text{rot}} \leq \frac{2}{3E}. \tag{7.103}$$

Now, $E = V\bar{d}$ where $\bar{d}$ is the mean degree of the network. Therefore Equation (7.102) requires:

$$\beta \bar{f}_{\text{rot}} \leq \frac{2}{3V\bar{d}}. \tag{7.104}$$

The original bound required that the largest rotational force was less than one over V times the maximum degree plus one half, while this bound requires that the mean rotational force is less than two over three times V times the mean degree.

Both bounds can be tighted using essentially the same algebra if we consider the power series with coefficients $||\hat{q}^{(k)}(\beta)||_2$. The first step is to bound $||\hat{q}(\beta)||_2$. This can be accomplished as follows. Symmetrize the weighted HHD so that it is of the form:

$$R^{-1/2}G\hat{q}^{(k)} + R^{1/2}C^{\intercal}\hat{\theta}^{(k)} = R^{-1/2}f^{(k)}.$$

Then, since $R^{-1/2}G$ is orthogonal to $R^{1/2}C^{\intercal}$:

$$||R^{-1/2}G\hat{q}^{(k)}||_2^2 + ||R^{1/2}C^{\intercal}\hat{\theta}^{(k)}||_2^2 = ||R^{-1/2}f^{(k)}||_2^2.$$

It follows that:

$$||R^{-1/2}G\hat{q}^{(k)}||_2^2 \leq ||R^{-1/2}f^{(k)}||_2^2. \tag{7.105}$$

Then, since $||R^{-1/2}G\hat{q}^{(k)}||_2^2 > \min_{ij}\{R_{ij}^{-1}\}||G\hat{q}^{(k)}||_2^2$ and $||R^{-1/2}f^{(k)}||_2^2 \leq ||R^{-1/2}||_2^2||f^{(k)}||_2^2$ which equals $\max_{ij}\{\tilde{R}_{ij}^{-1}\}||f^{(k)}||_2^2$:

$$||G\hat{q}^{(k)}||_2^2 \leq \frac{\min_{ij}\{R_{ij}\}}{\max_{ij}\{R_{ij}\}}||f^{(k)}||_2^2 = \kappa(R^{-1})||f^{(k)}||_2^2 \tag{7.106}$$

where $\kappa(R^{-1})$ is the condition number of the resistances.

The ratio:

$$\frac{||G\hat{q}^{(k)}||_2^2}{||\hat{q}^{(k)}||_2^2} = \frac{\hat{q}^{(k)\intercal}G^{\intercal}G\hat{q}^{(k)}}{\hat{q}^{(k)\intercal}\hat{q}^{(k)}} \tag{7.107}$$

is the Rayleigh quotient of a real symmetric matrix, so it is minimized by the smallest eigenvalue of $G^\intercal G$. We required that $\hat{q}^{(k)}$ had mean zero, so it has no projection along the null space of $G^\intercal G$. Therefore the quotient is minimized at smallest nonzero eigenvalue of the node Laplacian. The node Laplacian, $G^\intercal G$, is symmetric so this eigenvalue equals the smallest (nonzero) singular value σ_{V-1} of the Laplacian. Therefore:

$$||\hat{q}^{(k)}||_2 \leq \sqrt{\frac{1}{\sigma_{V-1}}}||G\hat{q}^{(k)}||_2 \leq \sqrt{\frac{\kappa(R^{-1})}{\sigma_{V-1}}}||f^{(k)}||_2. \tag{7.108}$$

Then, plugging in for $f^{(k)}$ and using the triangle inequality we arrive at the recursion:

$$||\hat{q}^{(k)}||_2 \leq \sqrt{\frac{2d\kappa(R^{-1})}{\sigma_{V-1}}} \sum_{j=1}^{k} \frac{||f_{\text{rot}}||_\infty^j}{j!}||\hat{q}^{(k-j)}||_2. \tag{7.109}$$

This recursion is of exactly the same form as the recursion derived for the infinity norm, only now $x = \sqrt{\frac{2d\kappa(R^{-1})}{\sigma_{V-1}}}$. Therefore the associated power series converges if:

$$\beta||f_{\text{rot}}||_2 \leq \frac{1}{\sqrt{\frac{2d\kappa(R^{-1})}{\sigma_{V-1}}} + \frac{1}{2}}. \tag{7.110}$$

It follows that the radius of convergence of the power series is greater than or equal to:

$$R \geq \frac{1}{\sqrt{\frac{2d\kappa(R^{-1})}{\sigma_{V-1}}} + \frac{1}{2}}. \tag{7.111}$$

Equation (7.111) is a significantly more complicated bound that balances the variation in the resistances, measured through the condition number of R, with the maximum degree and smallest singular value of $G^\intercal G$. If the variation in the resistances is small then this bound has some advantages. First, it depends on the square root of the maximum degree, not on the maximum degree. Second, in some important cases $\sqrt{1/\sigma_{V-1}}$ is much smaller

than V, in which case this bound may be much larger than the bound derived using either the infinity norm or the one norm.

For example, suppose that the network is an n dimensional lattice with side lengths $a_1m, a_2m, ...a_nm$, where $a_1 > a_2 > ...a_n$. Then $V = [\prod_{j=1}^{n} a_j]m^n$ and $d = 2^n$. Then the smallest nonzero singular value is (see Section 3.3.3):

$$\sigma_{V-1}(m) = 4\sin^2\left(\frac{\pi}{2a_1m}\right) \simeq \left(\frac{\pi}{a_1m}\right)^2.$$

For large m the bound using the infinity norm is:

$$\beta||f_{\text{rot}}||_\infty \leq \frac{1}{dV+1} \simeq \frac{1}{[\prod_{j=1}^{n} a_j](2m)^n}$$

which implies that the weak rotation regime is vanishing in $(2m)^n$, which implies that if m is large, then the weak rotation regime is very small. However, using the two norm bound:

$$\beta||f_{\text{rot}}||_\infty \leq \beta||f_{\text{rot}}||_2 \leq \frac{1}{\sqrt{\frac{2d\kappa(R^{-1})}{\sigma_{V-1}}}+1} \simeq \frac{\pi}{\sqrt{\kappa(R^{-1})}a_1m2^{n/2}}.$$

It follows that the weak rotation regime only vanishes proportional to $m2^{n/2}$. Therefore, as m goes to infinity the weak rotation regime may be vanishes no slower than $1/m$. This bound also reduces the rate at which the weak rotation regime vanishes in n to $2^{-n/2}$ instead of $(2m)^{-n}$. Both of these are huge improvements in the bounds. Suppose $n = 4$ and $m = 10^2$, as might be the case for a model representing a system of 4 competing species each with populations ranging from 1 to 100. Then the original bound requires that the rotation is weaker than $(1/16) \times 10^{-8}$ while the new bound requires that rotation is weaker than $\pi\kappa(R^{-1})^{-1/2} \times (1/4) \times 10^{-2}$. For moderate $\kappa(R^{-1})$ this is a much larger upper bound, so the weak rotation regime is orders of magnitude larger than suggested by our original

bound.

While Equation (7.111) is a significant improvement on the original bound, it still wildly underestimates the radius of convergence for real networks. For example, suppose we start with a hypercube and randomly prune a fixed fraction of the nodes. Then pick the largest connected component and study the convergence of the weak rotation expansion on that network.

Results for a sample network are described below. Pruning 40 percent of the nodes of a 7 dimensional hypercube left a connected component with 76 nodes, 152 edges, and 77 loops. Using randomly sampled conductances, ϕ and θ the square root condition number of the inverse resistances matrix R^{-1} was less than 151.7. The smallest nonzero singular value of the node Laplacian was 0.57^2 and the maximum degree was 7. Therefore the radius of convergence was, at least:

$$R > \frac{\pi}{151.7}\frac{0.57}{\sqrt{14}} \approx 0.003. \tag{7.112}$$

In practice the expansion was observed to converge for all $\beta||f_{\text{rot}}||_\infty < 1$. The same experiment was repeated 1000 times for percolation networks built by randomly removing nodes from high dimensional hypercubes, and in each case convergence was observed at the expected rates up to $\beta||f_{\text{rot}}||_\infty < 1$. This numerical observation inspires the conjecture that the weak rotation expansion converges for any $||f_{\text{rot}}||_\infty < 1$.

Rotation Independent Steady States

So far we have characterized the steady state for arbitrary ϕ, ρ and weak θ. We showed that the general problem of solving for the steady state can always be reduced to finding the steady state for a purely rotational process, and derived the exact steady state for an isolated loop or pair of loops. Here we seek conditions under which the steady state is

independent of f_{rot}. In particular, if $q(\beta)$ is constant in β with rotational forces βf_{rot} then the steady state is independent of f_{rot}. Since $q(\beta)$ is always uniform if $\beta = 0$ we conclude that $q(\beta)$ is independent if and only if it is uniform for all β. This extends nicely to the case when the purely rotational process is derived from a rescaling of a process that is not purely rotational. If the steady state of the purely rotational process is independent of f_{rot} then it is uniform, so the steady state of the full process is equal to the equilibrium distribution for the corresponding process in detailed balance.

Before launching into a general analysis we will consider a couple simple cases that will establish our expectations for the general theory.

First consider the case when the resistances R_{ij} are equal to a constant r. Then the first order correction to the steady state vanishes, and the first order correction to the fluxes is $R^{-1} f_{\text{rot}}$. What about the higher order corrections?

The second order equation is:

$$G\hat{q}^{(2)} + RC^{\intercal}\tilde{\theta}^{(2)} = F_{\text{rot}} H \hat{q}^{(1)} - \frac{1}{2} F_{\text{rot}}^2 G \hat{q}^{(0)}.$$

If R_{ij} are constant then $\hat{q}^{(1)} = 0$ so the first term on the right hand side vanishes. Moreover, $\hat{q}^{(0)}$ is always constant so $G\hat{q}^{(0)} = 0$. Therefore the right hand side is zero and $\hat{q}^{(2)} = 0, \tilde{\theta}^{(2)} = 0$. Therefore the steady state distribution is the equilibrium distribution to third order in ϵ, and the steady state currents are $\beta R^{-1} f_{\text{rot}}$ to third order in β.

The third order equation is:

$$G\hat{q}^{(2)} + RC^{\intercal}\tilde{\theta}^{(2)} = F_{\text{rot}} H \hat{q}^{(2)} - \frac{1}{2} F_{\text{rot}}^2 G \hat{q}^{(1)} + \frac{1}{6} F_{\text{rot}}^3 H \hat{q}^{(0)} = \frac{1}{3V} f_{\text{rot}}^3.$$

Even though f_{rot} is divergence free there is no guarantee that f_{rot}^3 is divergence free. For example, if we consider two loops sharing an edge, with $\theta = 1$ on the first and $\theta = 2$ on

the second, then the divergence of $\frac{1}{3v} f^3_{\text{rot}}$ is $+2/v$ and $-2/V$ at the junction nodes on either end of the edge where the loops meet. It follows that $\hat{q}^{(3)}$ is not necessarily zero.

Next consider k-connected components. A k-connected component of a graph is a connected subgraph that can be separated from the rest of the network by removing k edges [34]. At steady state the total flux into and out of any set of nodes must be equal to zero. Therefore, at steady state the total flux into any k-connected component must equal the total flux out. If k is small, say two or three, then this flux balance is easy to write and solve. This leads to simple rules for k connected components when k is small.

First, suppose $k = 1$. Then there is only one edge to consider. If $k = 1$ then there is no loop that includes the connecting edge so f_{rot} is zero on the edge. Therefore, the flux balance requires:

$$\rho_{ij}(q_i - q_j) = 0$$

where q_i, q_j are the steady state probabilities on either side of the edge. Notice that this requires $q_i = q_j$, so any 1-connected component always has uniform steady state probability on either end of the edge.

Now let $k = 2$. Then there are two edges connecting the component to the rest of the network. Number these edges 1 and 2. Since no loop can use the same edge twice, any loop passing over one of the edges must pass over the other in order to escape the component. This means that f_{rot} is the same on both edges. Let a be the shared rotational force. If the steady state is uniform then $q_i = 1/V$ at all nodes. Then, the flux balance for the component reads:

$$\frac{1}{V}(\rho_1(\exp(a) - \exp(-a) - \rho_2(\exp(a) - \exp(-a)) = \frac{\sinh(a)}{V}(\rho_1 - \rho_2) = 0$$

which is only zero if $\rho_1 = \rho_2$. Therefore, for any biconnected component, the steady state is

independent of rotation only if the conductances on the two connecting edges are identical. This rule is easy to apply, and can be used to exclude many simple examples from the set of networks whose steady state is independent of rotation. This is a familiar requirement, since it corresponds to avoiding bottlenecks in the first order corrections. A biconnected component is illustrated in Figure 7.7.

Now consider a triconnected component. Then there are, at most three different values of f_{rot} on the three edges connecting the component to the rest of the network. Label these f_1, f_2, f_3. These are in the span of the curl transpose, so can be rewritten in terms of a rotational potential. Any loop entering the component must leave on one of the other edges, so there are at most three unique ways for a loop to pass through the component. Either the loop uses edges 1 and 2, 2 and 3, or 3 and 1. Therefore the action of the curl transpose on the three edges is always of the form $f_1 = a - b$, $f_2 = b - c$, $f_3 = c - a$ for some values a, b, c. A triconnected component is illustrated in Figure 7.7. Now the flux balance equation has the form:

$$\frac{1}{V}[\rho_1 \sinh(a-b) + \rho_2 \sinh(b-c) + \rho_3 \sinh(c-a)] = 0.$$

Without loss of generality assume that $a \geq b \geq c$ so that the first two terms are nonnegative, and the last term is non-positive [3]. Clearly this equation is independent of V, so can be rewritten:

$$\rho_1 \sinh(a-b) + \rho_2 \sinh(b-c) + \rho_3 \sinh(c-a) = 0$$

[3] This is the general case because we can always reorder the edges/loops to make sure the vector potential runs in decreasing order as the edge index increases

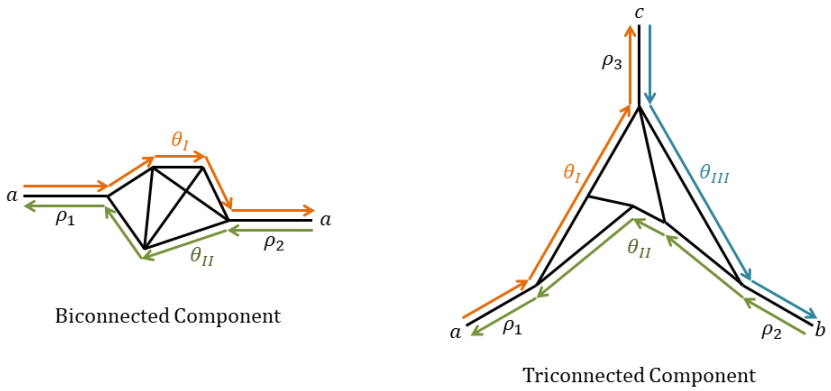

Figure 7.7: A biconnected and triconnected component. The rotational flow on the connecting edges is represented with arrows. The conductances on each of the connecting edges is labelled, and the value of the roational flow is marked with a, b, and c.

which is satisfied on the intersection of the plane:

$$\rho_3(\rho_1, \rho_2, a, b, c) = \frac{\sinh(a-b)}{\sinh(a-c)}\rho_1 + \frac{\sinh(b-c)}{\sinh(c-a)}\rho_2 \tag{7.113}$$

with the positive octant. This intersection is non-empty since both of the ratios of hyperbolic sines are positive, so, since ρ_1, ρ_2 are always greater than zero, $\rho_3 > 0$ can be chosen so that the steady state is uniform. When does the intersection of the plane with the positive octant include uniform weights?

If all the conductances are uniform then a, b, c must satisfy:

$$\sinh(a-b) + \sinh(b-c) + \sinh(c-a) = 0.$$

Let $x = a - b$ and $y = b - c$. Then $c - a = -(x + y)$ so we need:

$$\sinh x + \sinh y = \sinh(x + y).$$

This equation is only satisfied on the lines $x = 0$, $y = 0$ and $x = -y$.[4] These lines correspond to $a = b$, $b = c$ or $c = a$. But if any pair are equal then f_{rot} is zero on the corresponding edge, in which case the flux on the edge is automatically zero for a uniform distribution, and in which case the flux balance equation is the same as in the biconnected case. Therefore, if the network has any triconnected components, uniform weights on the three connected edges only leads to a uniform steady state if the rotational edge flow vanishes on at least one of the edges. Otherwise uniform weights never lead to a uniform steady state.

This analysis shows four important things. First, if there are any singularly connected components we can analyze each component separately since the flux over the connecting edge is automatically zero for a uniform distribution. Second, if the network contains biconnected components then it is only rotation independent if the resistances match on each pair of connecting edges. Third, if the network contains triconnected components then uniform resistances on the three connecting edges only lead to uniform steady state probabilities if the rotational flow is zero on one of the three edges. Fourth, a triconnected component may admit a uniform steady state if the resistances on the three connecting edges are chosen from the intersection of a plane with the positive quadrant, where the coefficients defining the plane depend on the rotational forces. This suggests that, while uniform resistances may not always lead to a uniform steady state, there may never the less be a space of resistances and rotational forces that admit a uniform steady state regardless the size of the rotational edge flow.

In general, for a uniform steady state the flux balance equation reads:

$$G^{\intercal} R^{-1} \sinh\left(\beta f_{\text{rot}}\right) = 0. \tag{7.114}$$

[4] Taylor expand both sides to third order, then the first order and second order terms vanish leaving $x^2 y = -y^2 x$. This gives the lines, which satisfy the exact equation.

Rotation independence requires that this equation is satisfied for all β. This is equivalent to requiring that all terms in the Taylor expansion of the left hand side are independently zero since the Taylor expansion of $\sinh$ converges everywhere. Then rotation independence requires:

$$G^\intercal R^{-1} f_{\text{rot}}^{2k+1} = 0 \tag{7.115}$$

for all k.

It is not obvious that this can be satisfied for arbitrary k, let alone all k up to a given order. More precisely, we would like to know to how many orders in β the steady state is independent of f_{rot}, that is, to what order the flux balance equation satisfied.

Consider the recursive equation for the Taylor expansion of the steady state:

$$-G^\intercal R^{-1} G q^{(k)} = G^\intercal R^{-1} \sum_{j=1}^{k} \frac{1}{j!} F_{\text{rot}}^j K_j q^{(k-j)}.$$

Since we require that the overall distribution is normalized all $q^{(k)}$ for $k > 1$ are orthogonal to the one dimensional nullspace of $G^\intercal W G$. Therefore each correction $q^{(k)} = 0$ if and only if the right hand side of the equation is zero.

For the first order term, setting the right hand side to zero requires:

$$G^\intercal R^{-1} f_{\text{rot}} = 0. \tag{7.116}$$

Fixing R^{-1} defines a linear subspace of f_{rot} where the steady state is rotation independent to first order, and fixing f_{rot} defines a linear subspace of resistances where the the steady state is rotation independent to first order.

In order for the steady state to be second order in β the first order equation must be

satisfied. Suppose $q^{(1)}$ is zero. Then right hand side of the second order equation reads:

$$\frac{1}{2}G^\intercal R^{-1}F_{\text{rot}}^2 G q^{(0)} \tag{7.117}$$

But $q^{(0)}$ is uniform so $Gq^{(0)} = 0$. Therefore, if the first order condition is satisfied, the second order correction is also zero.

Then, assuming independence up to second order, third order independence requires:

$$G^\intercal R^{-1}F_{\text{rot}}^3 H q^{(0)} = \frac{1}{3}G^\intercal R^{-1}f_{\text{rot}}^3 = 0. \tag{7.118}$$

The same process can be continued to arbitrary order. Then all terms on the right hand side that depend on corrections of order one or higher are assumed to vanish, so the only remaining term is $G^\intercal R^{-1}F_{\text{rot}}^j K_j q^{(0)}$. When j is even $K_j = G$, so the right hand side is automatically zero if the steady state is independent of rotation up to order $j-1$. If $j = 2k+1$ is odd then rotation independence requires:

$$G^\intercal R^{-1}f_{\text{rot}}^{2k+1} = 0. \tag{7.119}$$

This is exactly the condition we derived from the flux balance, but we can now relate the steady state to the largest k such that $G^\intercal R^{-1}f_{\text{rot}}^{2k+1} = 0$.

Lemma 33. *If $G^\intercal R^{-1}f_{rot}^{2k+1} = 0$ for all $k \leq n$ then the steady state is independent of rotation to order $2n+3$: $q(\beta) = \mathcal{O}(\beta^{2n+3})$, and the steady state is independent of rotation if and only if $G^\intercal R^{-1}f_{rot}^{2k+1} = 0$ for all k.*

Lemma 33 provides a simple criteria for testing rotation independence. Can we use this criteria to produce examples where the steady state is independent of rotation?

Suppose that f_{rot} is zero, one, or negative one on all edges. Then $f_{\text{rot}}^{2k+1} = f_{\text{rot}}$ so, if the resistances are chosen so that $G^\intercal R^{-1} f_{\text{rot}} = 0$ then the steady state is independent of rotation at all orders. This requires $R \in \text{null}\{G^\intercal F_{\text{rot}}\} \cap \mathbb{R}^{E+}$ where $\mathbb{R}^{E+}$ is the positive quadrant of the space of edges. Since this case reduces to the first order case set all of the resistances equal to a constant. Then the steady state is independent of rotation to all orders.

A more general analysis begins by fixing f_{rot}. Then the steady state is rotation independent up to order $2n+3$ if for all $k \leq n$:

$$G^\intercal R^{-1} f_{\text{rot}}^{2k+1} = 0.$$

Let:

$$F^{(n)} = \begin{bmatrix} f_{rot_1} & f_{rot_2} & \cdots & f_{rot_E} \\ f_{rot_1}^3 & f_{rot_2}^3 & \cdots & f_{rot_E}^3 \\ \vdots & \vdots & \ddots & \vdots \\ f_{rot_1}^{2n+1} & f_{rot_2}^{2n+1} & \cdots & f_{rot_E}^{2n+1} \end{bmatrix}. \tag{7.120}$$

Then the steady state is rotation independent up to order $2n+3$ if:

$$F^{(n)} R^{-1} G = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \tag{7.121}$$

where the matrix of zeros on the right hand side is n by V. Equation (7.121) requires $\text{range}\{G\} \subseteq \text{null}\{F^{(n)} R^{-1}\}$. The dimension of the range of G is always $V-1$, so this is only possible if the dimension of the nullspace of $F^{(n)} W$ is greater than or equal to $V-1$.

The resistances are all nonzero, so the diagonal matrix R^{-1} is invertible. Therefore the dimension of the null$\{F^{(n)}R^{-1}\}$ is the same as |null$\{F^{(n)}\}$|. Therefore, if |null$\{F^{(n)}W\}|<$ $V-1$ then the steady state $q(\epsilon)$ is, at most, independent of rotation to order $2n+1$. That is $q(\beta) \geq \mathcal{O}(\beta^{2n+1})$.

Thus we can bound the order to which the steady state is independent of rotation simply by studying the nullity of $F^{(n)}$. The matrix $F^{(n)}$ is similar to a Vandermonde matrix [51, 263]. Define the Vandermonde matrix:

$$V^{(n)} = \begin{bmatrix} 1 & f_{rot_1}^2 & \dots & f_{rot_1}^{2n} \\ 1 & f_{rot_2}^2 & \dots & f_{rot_2}^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_{rot_E}^2 & \dots & f_{rot_E}^{2n} \end{bmatrix} \tag{7.122}$$

then:

$$F^{(n)} = V^{(n)\mathsf{T}} F_{\text{rot}}. \tag{7.123}$$

The rank and nullity of Vandermonde matrices are well characterized since Vandermonde matrices are important in polynomial fitting [51, 264]. To take advantage of the Vandermonde matrix, rephrase the problem in terms of the rank of $F^{(n)}$. $F^{(n)}$ is an $n \times E$ matrix so:

$$|\text{null}\{F^{(n)}\}| = E - \text{rank}\left(F^{(n)\mathsf{T}}\right) = E - \text{rank}(F_{\text{rot}} V^{(n)}). \tag{7.124}$$

Next note that removing rows from F_{rot} and $V^{(n)}$ that correspond to edges with rotational force zero does not change the rank of the matrix $F_{\text{rot}}V^{(n)}$ since a row of all zeros produces a zero entry no matter what the matrix is multiplied with. Also notice that if f_{rot}^2 takes on the same value on some subset of the edges then the corresponding subset of

rows of $F_{\text{rot}}V^{(n)}$ will all be identical. A series of redundant rows can be replaced with a single row containing the same values without changing the rank, since these rows are not independent.

Define $\hat{V}^{(n)}$ and $\hat{F_{\text{rot}}}$ to be $V^{(n)}$ and F_{rot} with all zero rows removed, and duplicated rows replaced with a single row. Let m be the number of distinct nonzero values taken on by $f^2_{rot_{ij}}$. Then $\hat{F_{\text{rot}}}$ is a diagonal $m \times m$ matrix with nonzero entries, so it is invertible and does not change the rank. Therefore:

$$|\text{null}\{F^{(n)}\}| = E - \text{rank}(\hat{V}^{(n)}) \tag{7.125}$$

where $\hat{V}^{(n)}$ is a $m \times n$ Vandermonde matrix with distinct rows. The rank of a $m \times n$ Vandermonde matrix with all distinct rows is $\min\{m, n\}$ therefore:

$$|\text{null}\{F^{(n)}\}| = E - \min\{m, n\} \tag{7.126}$$

Therefore:

Lemma 34. *Let E be the number of edges, V the number of nodes, and $L = E - (V - 1)$ the number of loops in the network. Let m be the number of distinct nonzero values of f^2_{rot}. If $m > L$, then the steady state cannot be independent of rotation to all orders and, at most $q(\beta)$ is $\mathcal{O}(\beta^{2L+1})$. If $m \leq L$ then it is possible that $q(\beta)$ is independent of f_{rot} to all orders.*

Lemma 34 provides an upper bound on the number of distinct values of f^2_{rot}, which, if exceeded, guarantees that the steady state is never rotation independent. In that case the steady state is never independent of rotation to order higher than $2L + 1$. Alternatively, if there are sufficiently few distinct nonzero values of f^2_{rot} then the conditions $G^\intercal R^{-1} f^{2k+1}_{\text{rot}}$ can become redundant for large enough k, so it is possible that the steady state is entirely

independent of rotation.

As a corollary, if any two loops overlap then it is possible that there are more distinct nonzero values of f_{rot}^2 than there are loops in the network. In that case it is always possible that $m > L$. If $m > L$ then the steady state is never independent of rotation to all orders. Therefore, if the loops in the network are not disjoint, there is never a set of resistances that guarantee rotation independence for all f_{rot}. This rules out the possibility that uniform resistances always lead to rotation independence. In short, there is no rule for generating resistances that guarantee rotation independence. Instead, for some f_{rot} there is a space of resistances that can guarantee independence. The next lemma addresses the dimension of this space.

Fix f_{rot}. Let $w_{ij} = 1/R_{ij}$ be a set of weights corresponding to the reciprocal of the resistances. In order for the steady state to be independent of rotation up to order $2n + 1$ the weights must satisfy the linear equations:

$$\begin{bmatrix} G^\intercal F_{\text{rot}} \\ G^\intercal F_{\text{rot}}^3 \\ \vdots \\ G^\intercal F_{\text{rot}}^{2n+1} \end{bmatrix} w = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{7.127}$$

Therefore, the space of weights such that the steady state is independent of the rotational forces up to order n is the intersection of $\mathbb{R}^{E+}$ with a linear subspace:

$$\{w | q(\epsilon) \text{ is } \mathcal{O}(\epsilon^{2n+3})\} = \text{null}\left\{ \begin{bmatrix} GF_{\text{rot}} & GF_{\text{rot}}^3 & \dots & GF_{\text{rot}}^{2n+1} \end{bmatrix}^\intercal \right\} \cap \mathbb{R}^{E+}$$

These spaces are nested, since at each order we introduce new conditions (columns to the block matrix) that w must satisfy. It is natural to ask, for a given n, what is the dimension

of this subspace of weights? Does it always decrease when n increases? If so, how fast? Using the fundamental theorem of linear algebra [26]:

$$|\{w|q(\epsilon) \text{ is } \mathcal{O}(\epsilon^{2n+3})\}| = E - \text{rank}\left(\left[\begin{array}{cccc} F_{\text{rot}}G & F_{\text{rot}}^3 G & \dots & F_{\text{rot}}^{2n+1}G \end{array}\right]\right).$$

The matrix $\left[\begin{array}{cccc} F_{\text{rot}}G & F_{\text{rot}}^3 G & \dots & F_{\text{rot}}^{2n+1}G \end{array}\right]$ is $E \times nV$. Each block has, at most, rank $V-1$ since G has rank $V-1$. At most, all the blocks are independent, in which case the rank of the whole matrix is always less than $\min\{E, n(V-1)\}$. Therefore:

$$|\{w|q(\epsilon) \text{ is } \mathcal{O}(\epsilon^{2n+3})\}| \geq E - n(V-1) \tag{7.128}$$

if it is non-empty.

This gives a lower bound on the dimension of the space of weights such that the steady state is independent of rotation up to order n. This bound can be tightened if f_{rot}^2 takes on only a few nonzero values.

Let B be a basis for the range of G that is full (all its entries are nonzero). Now, let b_j be the j^{th} column of B. Then we can replace G with B without changing the rank of the block matrix. Next, rearrange the columns of the block matrix so that it has the form:

$$\left[\begin{array}{cccc} F_{\text{rot}}K_n(F_{\text{rot}}^2, b_1) & F_{\text{rot}}K_n(F_{\text{rot}}^2, b_2) & \dots & F_{\text{rot}}K_n(F_{\text{rot}}^2, b_{V-1}) \end{array}\right].$$

where $K_n(A, b)$ is the Krylov matrix:

$$\left[\begin{array}{ccccc} b & Ab & A^2 b & \dots & A^n b \end{array}\right]. \tag{7.129}$$

Permuting columns does not change the rank.

At most, each of the $V-1$ blocks are linearly independent so:

$$\text{rank}\left(\left[\begin{array}{cccc} F_{\text{rot}}G & F_{\text{rot}}^3G & \dots & F_{\text{rot}}^{2n+1}G \end{array}\right]\right) \leq \sum_{j=1}^{V-1} \text{rank}(F_{\text{rot}}K_n(F_{\text{rot}}^2, b_j)).$$

The j^{th} block has form:

$$F_{\text{rot}}K_n(F_{\text{rot}}^2, b_j) = F_{\text{rot}}V^{(n)}\text{diag}(b_j). \tag{7.130}$$

Since B was chosen so that it was full, $\text{diag}(b_j)$ is invertible and does not change the rank. The rank of $F_{\text{rot}}V^{(n)}$ is $\min\{m,n\}$. This bound is independent of j so:

$$\text{rank}\left(\left[\begin{array}{cccc} F_{\text{rot}}G & F_{\text{rot}}^3G & \dots & F_{\text{rot}}^{2n+1}G \end{array}\right]\right) \leq \min\{m,n\}(V-1). \tag{7.131}$$

Therefore:

Lemma 35. *The space of weights $w = 1/R$ such that the steady state is independent of rotation up to order $2n+3$ is the intersection of a linear subspace with $\mathbb{R}^{E+}$. This defines a space that is either empty (does not intersect the $\mathbb{R}^{E+}$), or has dimension equal to the dimension of the subspace. Therefore, the space of weights such that the steady state is independent of rotation up to order $2n+3$ has dimension:*

$$|\{w|q(\beta) \textit{ is } \mathcal{O}(\beta^{2n+3})\}| \geq E - \min\{m,n\}(V-1) \tag{7.132}$$

if it is non-empty. If $m < \lfloor \frac{E}{V-1} \rfloor$ then the space of weights such that $q(\beta)$ is independent of rotation to all orders has dimension $E - m(V-1)$ if it non-empty.

To summarize, Lemma 33 establishes the exact condition for rotation independence,

Lemma 34 gives an upper bound on the number of distinct nonzero values of f_{rot}^2, m, past which the steady state is never rotation independent, and Lemma 35 gives a threshold on m, where, if m is less than the threshold, then the dimension of space of weights such that the steady state is rotation independent has a nonzero lower bound. Therefore the number of distinct nonzero values of f_{rot}^2 is the essential number, which, if sufficiently small, allows for rotation independent steady states.

Dependence on m may seem a little strange, but it is natural given the nature of the flux balance conditions. Each order only differs by changing the power of f_{rot} used. Every time we increase the order we add a new set of equations equivalent to a block of the previous equations scaled by an additional factor of f_{rot}^2. If a block of equations is independent of the previous block then the space of possible weights/resistances shrinks every time we increase the order of independence. Since the space is finite, if enough independent blocks are added then the space of weights vanishes. However, if f_{rot}^2 only takes on m distinct nonzero values then multiplication by f_{rot}^2 only produces new equations up to m times. Therefore we have at most $\min\{m, n\}(V-1)$ linearly independent equations. For $n > m$ adding higher orders does not change the space, so, if m is sufficiently small we can guarantee that the space is nonempty, and of a certain minimal size. Alternatively, if m is too large then the space of weights such that the steady state is rotation independent to order $2n+1$ keeps shrinking each time n increases until it is empty. This takes, at most, $2L+1$ steps, so the steady state is, at most, independent of rotation to order $2L+1$.

7.3.3 Strong Forcing Limits

In a strong forcing limit the components of the edge flow, f are all assumed to be large. If the components of the edge flow are large then the forward transition rates on each edge are much faster than the backward transition rates. As a consequence, the sequence of

states visited becomes highly directed, and, depending how the limit is taken, increasingly predictable. A strong forcing limit is achieved by scaling the edge flow by a large number. Here we consider three different strong forcing limits, since the behavior of the steady state depends on how the limit is performed. In all cases the steady state is described by an object that mimics a potential constructed by evaluating the work over some set of optimal paths. The set of paths, and the definition of optimal depends on the limit taken. These are analogous to the quasipotential used to analyze SDE's in the small noise limit. If the noise is sufficiently small then the steady state of an SDE is described by a quasipotential, which is defined by evaluating the work required to reach each point in space from a stable equilibrium of the deterministic process along the most likely trajectory from the equilibrium [23].

Strong Forcing

As when considering weak forcing limits, start by introducing a scaling β that is analogous to inverse temperature, and where the transition rates are parameterized by:

$$l_{ij}(\beta) = \rho_{ij} \exp(\beta f_{ij}).$$

In the strong forcing limit β diverges to $+\infty$ instead of 0.

Notice that if $f_{ij} > 0$ then $l_{ij}(\beta)$ diverges as β diverges, and, since $f_{ji} = -f_{ij}$, $l_{ji}(\beta)$ converges to zero as β diverges. Therefore, in the strong forcing limit, the ratio of the forward to backward transition rate diverges on every edge. This makes expansion of $L(\beta)$ in β difficult. Troublingly, if $f_{ij} > f_{kh}$ then the ratio $l_{kh}(\beta)/l_{ij}(\beta)$ converges to zero as β diverges, even if $l_{kh}(\beta)$ diverges as β diverges. Not only do half the nonzero elements of $L(\beta)$ diverge, they all diverge at asymptotically different rates.

For these reasons we do not attempt to solve for the stationary distribution by expanding $L(\beta)$. Instead we start with an exact expression for the stationary distribution in terms of the transition rates, and consider its limit. The following construction was borrowed from [5].

Denote the set of spanning trees of the network $\mathcal{T}$. Index all the distinct spanning trees in $\mathcal{T}$ so that the index ν corresponds to a specific tree: $T_\nu \in \mathcal{T}$. Pick a spanning tree T_ν. Now pick a specific node i. Define the directed tree $T_\nu(i)$ that consists of all directed edges whose undirected edges are elements of T_ν, and point along the paths in T_ν back to i. That is, $T_\nu(i)$ is the union of all paths from the leaves of T_ν to i, along the branches of T_ν. Index all the undirected edges in the graph. Let K_ν be the set of edge indices corresponding to edges in T_ν. Let $s(k|\nu, i)$ be 1 if the directed edge in $T_\nu(i)$ corresponding to undirected edge k points in the forward direction, and -1 if it points in the reverse direction.

Then, the stationary distribution at node i is [5]:

$$q_i(\beta) = \frac{1}{Z(\beta)} \sum_\nu \left(\prod_{k \in K_\nu} \rho_k \right) \exp\left(\sum_{k \in K_\nu} \beta s(k|\nu, i) f_k \right) \tag{7.133}$$

where $Z(\beta)$ is the necessary normalizing constant.

Let $\rho(\nu)$ be the total conductance of the tree T_ν:

$$\rho(\nu) = \prod_{k \in K_\nu} \rho_k \tag{7.134}$$

and let $w(\nu, i)$ denote the total work needed to move from the leaves of the directed tree to node i:

$$w(\nu, i) = \sum_{k \in K_\nu} s(k|\nu, i) f_k. \tag{7.135}$$

Then:

$$q_i(\beta) = \frac{1}{Z(\beta)} \sum_{\nu} \rho(\nu) \exp\left(\beta w(\nu, i)\right). \tag{7.136}$$

To avoid working with the limit of the partition function pick a reference node j and measure the stationary distribution at i relative to the distribution at j. Then:

$$\frac{q_i(\beta)}{q_j(\beta)} = \frac{\sum_{\nu} \rho(\nu) \exp\left(\beta w(\nu, i)\right)}{\sum_{\nu} \rho(\nu) \exp\left(\beta w(\nu, j)\right)} \tag{7.137}$$

Next consider the limit as β goes to infinity. Provided $|w(\nu, i)| > 0$ the exponential term $\exp\left(\beta w(\nu, i)\right)$ either converges to zero or diverges. Depending on the value of $w(\nu, i)$ the exponential diverges to infinity or converges to zero at different asymptotic rates. Therefore the sum over $\exp\left(\beta w(\nu, i)\right)$ will be dominated by the tree T_ν that maximizes $w(\nu, i)$. Define:

$$\begin{aligned} W(i) &= \max_{\nu} \{w(\nu, i)\} \\ M(i) &= \text{argmax}_{\nu} \{w(\nu, i)\}. \end{aligned} \tag{7.138}$$

Note that $M(i)$ is the set of trees which maximize $w(\nu, i)$, so may include more than one tree. Also note that, since the path dependent part of the work to move between two points only depends on f_{rot}, the optimal spanning trees are determined exclusively by the rotational part of the edge flow.

Then:

$$\begin{aligned} &\sum_{\nu} \rho(\nu) \exp\left(\beta w(\nu, i)\right) \\ &\qquad = \exp\left(\beta W(i)\right) \left[\sum_{\nu \in M(i)} \rho(\nu) + \sum_{\nu \notin M(i)} \rho(\nu) \exp\left(\beta (w(\nu, i) - W(i))\right) \right]. \end{aligned} \tag{7.139}$$

Since $W(i)$ maximizes $w(\nu, i)$ the last term, $(w(\nu, i) - W(i))$, is strictly negative.

Therefore $\rho(\nu)\exp\left(\beta(w(\nu,i)-W(i))\right)$ vanishes exponentially fast as β goes to infinity. Define:

$$\epsilon_i(\beta) = \frac{1}{\sum_{\nu\in M(i)}\rho(\nu)} \sum_{\nu\notin M(i)} \rho(\nu)\exp\left(\beta(w(\nu,i)-W(i))\right). \tag{7.140}$$

Then:

$$\frac{q_i(\beta)}{q_j(\beta)} = \exp\left(\beta(W(i)-W(j))\right)\frac{\sum_{\nu\in M(i)}\rho(\nu)}{\sum_{\nu\in M(j)}\rho(\nu)}\left[\frac{1+\epsilon_i(\beta)}{1+\epsilon_j(\beta)}\right]. \tag{7.141}$$

To study the limiting behavior of the steady state consider the effective potential. The effective potential is the negative log of the steady state distribution, scaled by the inverse temperature:

$$\phi_{eff}(\beta) = -\frac{1}{\beta}\log\left(q(\beta)\right) \tag{7.142}$$

The effective potential would equal the scalar potential if the system obeyed detailed balance.

Then, the difference in effective potential at a pair of nodes is the log-ratio of the steady states:

$$\phi_{eff_j}(\beta) - \phi_{eff_i}(\beta) = \frac{1}{\beta}\log\left(\frac{q_i(\beta)}{q_j(\beta)}\right).$$

Substituting Equation (7.141) for the ratio of the steady states yields:

$$\begin{aligned}\phi_{eff_j}(\beta) - \phi_{eff_i}(\beta) = &(W(i)-W(j)) + \frac{1}{\beta}\log\left(\frac{\sum_{\nu\in M(i)}\rho(\nu)}{\sum_{\nu\in M(j)}\rho(\nu)}\right)\\ &+\frac{1}{\beta}\left[\log\left(1+\epsilon_i(\beta)\right)-\log\left(1+\epsilon_j(\beta)\right)\right].\end{aligned}$$

As β diverges $\epsilon(\beta)$ converge to zero exponentially, so:

$$\lim_{\beta\to\infty}\left[\log\left(1+\epsilon_i(\beta)\right)-\log\left(1+\epsilon_j(\beta)\right)\right] = \lim_{\beta\to\infty}\epsilon_i(\beta)-\epsilon_j(\beta).$$

The vanishing terms, $\epsilon(\beta)$, converge to zero exponentially fast as β diverges. Therefore: $\epsilon_i(\beta) - \epsilon_j(\beta)$ vanishes faster than $1/\beta$. So for large β:

$$[\phi_{eff_j}(\beta) - \phi_{eff_i}(\beta)] \cong (W(i) - W(j)) + \frac{1}{\beta} \log \left(\frac{\sum_{\nu \in M(i)} \rho(\nu)}{\sum_{\nu \in M(j)} \rho(\nu)} \right) + \mathcal{O}(\exp(-\beta)). \tag{7.143}$$

Equation (7.143) states that, in the strong forcing limit the difference in the effective potential (log of the steady state divided by the large parameter β) between two vertices converges to the difference in maximal work to move to each node on an optimal spanning tree. The difference in the two spanning trees defines an ensemble of trajectories from i to j. Thus, in the strong forcing limit the effective potential converges to an object that is similar, but not identical, to the quasipotential. Instead of evaluating the work over one optimal trajectory between two points we evaluate work over a set of optimal trajectories formed by the difference between two optimal spanning trees. An example is illustrated in Figure 7.8

Notice that $-W(j)$ is the work it takes to move away from node j and $W(i)$ is the work it takes to move to node i. If $M(i) = \nu = M(j)$ then both trees are the same. Therefore all the edges of the directed tree $T_\nu(i)$ and $T_\nu(j)$ are oriented in the same direction, except the edges on the path from j to i in T_ν. To see this fact separate the tree T_ν into three sets of edges. The first set is the path from i to j. This acts like a bridge. Imagine removing the bridge from the tree. Removing the bridge separates the tree into two or more disjoint components. The first component is a tree branching out from node i, and the second is a tree branching out from node j. Any additional components branch off of the path. Now, orient the edges in the first component to point to node i. Since the bridge connecting i and j is the only way to move from the first component to node j orienting the edges in the first component towards node i is the same as orienting the edges to node j. Furthermore any

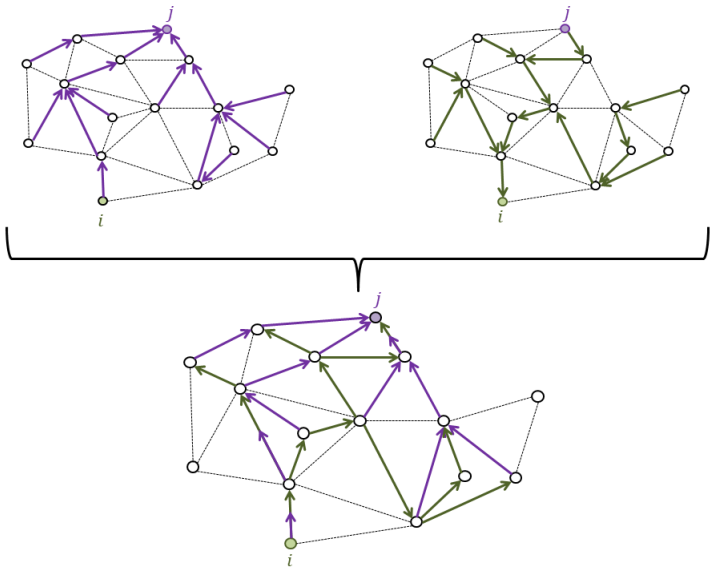

Figure 7.8: The difference between an optimal spanning tree directed towards node j (shown in purple in the upper left), and an optimal spanning tree directed towards node i (shown in green in the upper right), forms an ensemble of paths from i to j (the large network shown in the bottom center).

edges in components branching off the path must be oriented towards the path. Therefore the contribution to the work $W(i)$ from all edges not on the path from j to i is equal to the contribution to $W(j)$, so cancels out when taking $W(i) - W(j)$. This just leaves twice the work to move from j to i along the path in T_ν. So, in the special case when $M(i) = M(j)$ the difference in effective potential between i and j is twice the work it takes to move from i to j along a path in a shared optimal spanning tree.

It also follows that if there exist a set of nodes that can be separated from i and j by removing one edge then the trees associated with i and j must be identical on these nodes, so the steady state ratio between i and j is independent of the transition rates between those nodes.

In the next two sections we will develop an alternative procedure for analyzing strong forcing limit. This procedure is based on studying the dynamics of the skeleton process,

the sequence of states visited by the full process $X(t)$, and allows for a more general treatment of the conductances. This more general treatment includes limiting processes where the transition rates don't diverge, even if the ratio of forward and backward rates diverges. This generalization is important since the usefulness of a limit usually depends on how well it approximates examples that arise in applications, and the limit analyzed in this section produces strange Markov processes with time scale separation between all rates. It also leads to an alternative quasipotential-esque construction which is more useful for comparison to the quasipotential in the continuum.

Strong Rotation

How does the steady state behave when rotation is strong?

Here we consider three different strong rotation limits. In all cases the transition rates are of the form:

$$l_{ij}(\beta) = \rho_{ij}(\beta) \exp\left(\beta f_{\text{rot}_{ij}}\right)$$

where β is an parameter that is used to increase the strength of rotation. In applications β may be a function of physical parameters which influence the transition rates (for example, effective population size). Let $\alpha = \{\alpha_1, \alpha_2, ...\}$ be a sequence of parameters. Then suppose:

$$l_{ij}(\alpha) = \rho_{ij}(\alpha) \exp\left(\beta(\alpha) f_{\text{rot}}(\alpha)_{ij}\right)$$

and consider a limit in which $\beta(\alpha)$ diverges to infinity and $f_{\text{rot}}(\alpha)$ converges to a constant. The model and limiting value of α determine the (scaled) rotational flow, and whether the conductances $\rho_{ij}(\alpha)$ may converge or diverge. The generalized treatment in terms of β is motivated by assuming that we fix some sequence of parameters $\{\alpha_1, \alpha_2, ...\}$ so that β is diverging monotonically, and the scaled rotational forces approach their limit. Since $\beta(\alpha_j)$

is assumed to increase monotonically we use the shorthand $\rho_{ij}(\beta)$. This leaves the behavior of $\rho_{ij}(\beta)$ up to the parameter dependence and limiting sequence of parameters. Therefore, if we seek a general theory for strong rotation limits it is important to consider different limiting behavior in the conductances. We will organize our investigation by introducing three different constraints on the possible behavior of the conductances when β is large.

The three limits of interest correspond to different assumptions about the behavior of the conductances as β becomes large. The limits lead to three qualitatively different classes of Markov processes that can be approximated with a large rotation limit.

The three limits we consider are specified by the following three constraints:

1. The geometric mean of each pair of forward and backward transition rates ($\rho_{ij}(\beta)$) converge to finite nonzero constants in β.

2. The arithmetic mean of each pair of forward and backward transition rates converges to finite nonzero constants in β.

3. The expected waiting time in any node converges to a finite nonzero constant in β.

We will show that when β is large these processes converge to Markov chains with the following properties:

1. Under all three assumptions the ratio of forward to backward transition rates diverges on all edges where the rotational forces are not zero. This leads to processes where most, if not every, transition becomes close to irreversible.

2. Under the first or third constraint the neighbors of every node can be partitioned into two sets. The first set is the set of neighbors $i \in \mathcal{N}_j$ for which $f_{rot_{ij}}$ is maximized. The second is the complement of the first. Then the probability that we observe a

transition from j to i converges to zero if i is in the second set, and converges to a constant that only depends on $\rho_{ij}(\beta)$ if i is in the first set and β is large. If there is a unique neighbor that maximizes $f_{rot_{ij}}$ then the skeleton process converges to a deterministic process.

3. Under the first condition the rate of the forward transition diverges exponentially in β on all edges with nonzero rotational forces. This means that the process moves faster and faster as β increases. Moreover, provided f_{rot} varies in size across the network, the relative rate of transition on any pair of transitions for which f_{rot} differs will also diverge. This leads to a large time scale separation between all sets of transitions corresponding to different values of f_{rot}. This is the rotational limitation of the limit studied in Section 7.3.3. It is worth considering the purely rotational version of the strong forcing limit since any nonequilibrium Markov chain that satisfies microscopic reversibility can be transformed into a purely rotational process via the transform described in Section 7.2.1. Then, by studying the strong rotational limit independently we can find the limiting behavior of processes which are strongly rotational but could have a small conservative component.

Since we are interested in using these limits to approximate models arising from applications it is important to have this qualitative characterization to help pick which limit most closely matches the model of interest. These limits are appropriate when the model of interest is purely rotational, or has been transformed into a purely rotational model, and has close to irreversible transitions. The first limiting scenario is appropriate when the corresponding skeleton process is close to deterministic and there is large time scale separation between the rates that corresponds to the values of f_{rot}. The third limiting scenario is appropriate when the skeleton process is close to deterministic, but there is not

large time scale separation between the waiting times in each node. The second limiting scenario is appropriate when the skeleton process is not close to deterministic and the forward transition rates do not depend primarily on f_{rot}.

We will focus primarily on the first and third limits (the case when the skeleton process converges towards a deterministic process) since these offer the most useful comparisons to the quasipotential in the continuum, and are the most tractable. The second limiting case converges to a set of Markov processes that are not well described by the decomposition into conductances and edge flows, so are best treated with other methods. When the skeleton process convergences towards a deterministic process the asymptotic analysis is simpler.

Consider the first limiting scenario ($\rho_{ij}(\beta)$ converge to constants). Then the transition rates are approximated by $l_{ij}(\beta) = \rho_{ij} \exp(\beta f_{rot_{ij}})$ where ρ_{ij} and $f_{rot_{ij}}$ are set to their limiting values. Let $q(\beta)$ be the corresponding steady state distribution. We would like to approximate $q(\beta)$ when β is large. As before, finding the asymptotic behavior of $q(\beta)$ in a strong rotation limit is trickier than in the weak rotation limit (when β is small) because $l_{ij}(\beta)$ cannot be approximated with a Taylor expansion in ϵ with $\beta = 1/\epsilon$ or a Laurent expansion in β. As a consequence, the operator $L(\beta)$ cannot be trivially expanded to yield a recursive sequence of correction equations.

Here it is helpful to consider the skeleton process. The full process is the continuous time process $X(t)$. Note that any trajectory of the full process can be broken into a list of states $X = \{X_0, X_1, ...\}$ and a list of the time spent in each state $T = \{T_0, T_1, ...\}$. The waiting times T_j are independent of all X_i except X_j and are drawn from an exponential distribution with rate $\sum_{i \in \mathcal{N}_j} l_{ij}(\beta)$. The sequence of states is independent of the waiting times. It follows that the list of states can be drawn without drawing T. The skeleton process is the discrete time process corresponding to the sequence of states $X = X_0, X_1, ..,$

where time is recorded in the number of transition events.

The skeleton process is a discrete time Markov chain with transition probabilities:

$$\hat{l}_{ij}(\beta) = \frac{l_{ij}(\beta)}{\sum_{k \in \mathcal{N}_j} l_{kj}(\beta)}. \tag{7.144}$$

Let τ_j denote the expected waiting time to a transition out of node j. Then $\tau_j = \mathbb{E}[T_n | X_n = j] = \left(\sum_{k \in \mathcal{N}_j} l_{kj}(\beta)\right)^{-1}$. Note that, since we assumed the full process was connected and satisfied microscopic reversibility, the skeleton process is irreducible. The skeleton process may or may not be aperiodic, so that it may not have a unique steady state. If it is not aperiodic then the distribution converges to a periodic sequence of distributions whose phase depends on the initial condition. In that case define the steady state to be the average distribution across a full period. Let $\hat{q}(\beta)$ be the steady state of the skeleton process. The steady state of the full process can be recovered from the steady state of the skeleton process by:

$$q_i(\beta) = \frac{\tau_i(\beta)}{\tau(\beta)} \hat{q}_i(\beta) \tag{7.145}$$

where $\tau(\beta) = \sum_i \tau_i(\beta) \hat{q}_i(\beta)$ is the expected waiting time in any node at the steady state distribution. Equation (7.145) can be easily checked by plugging back into the steady state equation $Lq = 0$ since $l_{ij} q_j \propto \hat{l}_{ij} \hat{q}_j$.

Unlike the full process, whose forward transition rates diverge to infinity at asymptotically different rates in β, the transition probabilities of the skeleton process converge to finite values. Thus it is easier approximate $q(\beta)$ by estimating $\hat{q}(\beta)$ and then scaling by $\tau_i(\beta)$ than it is to approximate $q(\beta)$ directly.

For each node i define the maximal edge set:

$$M_i = \text{argmax}_{j \in \mathcal{N}_j} \{f_{rot_{ij}}\}. \tag{7.146}$$

Then let:

$$\Delta f_{ij} = f_{rot_{M_i}} - f_{rot_{ij}} \tag{7.147}$$

be the difference between the largest flow leaving node i and the flow on edge ij. For all $j \notin M_i$, $\Delta f_{ij} > 0$. Define the small quantity:

$$\epsilon_{ij}(\beta) = \frac{\rho_{ij}}{\rho_{M_i}} \exp\left(-\beta \Delta f_{ij}\right). \tag{7.148}$$

Since $\Delta f_{ij} > 0$ for all $j \notin M_i$, $\epsilon_{ij}(\beta)$ converges to zero exponentially fast in β. That is:

$$\epsilon_{ij}(\beta) = \mathcal{O}(\exp\left(-\beta \Delta f_{ij}\right)). \tag{7.149}$$

Now let:

$$\epsilon_i(\beta) = \sum_{j \notin M_i} \epsilon_{ij}(\beta). \tag{7.150}$$

and let:

$$\Delta f_i = \min_{j \notin M_i} \{\Delta f_{ij}\}. \tag{7.151}$$

Then, $\epsilon_i(\beta)$ is vanishing exponentially fast in β and:

$$\epsilon_i(\beta) = \mathcal{O}(\exp\left(-\beta \Delta f_i\right)). \tag{7.152}$$

The small parameters ϵ can be used to build an approximation to the steady state in the large β limit. Note that the small parameters are all exponential in $-\beta$ and the asymptotic rate at which they converge to zero depends on the difference between the largest and second largest rotational flow leaving each node.

In order to simplify the analysis we introduce two assumptions. These limit the scope

of the analysis, but are introduced to rule out edge cases. They are not overly limiting since they are each satisfied on an E dimensional space of possible flows and fail on lower dimensional subspaces of the set of possible flows. Therefore the two assumptions can be thought of as defining a general case, and fail in special cases when some of the edge flows leaving a node are exactly equal.

1. Assumption 1: For all i there exists an j such that $f_{rot_{ij}} \neq 0$.

2. Assumption 2: For each i there is a unique j that maximizes $f_{rot_{ij}}$. That is, $|M_i|= 1$ for all i.

The first assumption ensures that the skeleton process becomes irreversible in the large rotation limit. Combined the two assumptions ensure that the skeleton process converges to a deterministic process in the large rotation limit.

Since f_{rot} is purely rotational $G^\intercal f_{\text{rot}} = 0$. Therefore, the sum of f_{rot} on all edges surrounding any node is zero. It follows that, either f_{rot} is zero on all edges neighboring a particular node, or that $f_{\text{rot}} < 0$ on some of the edges and $f_{\text{rot}} > 0$ on some of the remaining edges. If we require that, for all i there exists an j such that $f_{rot_{ij}} \neq 0$, then the first case is impossible. Therefore the first assumption ensures that, for all i there is some edge ij such that $f_{\text{rot}_{ij}} > 0$. It follows that $f_{rot_{M_i}} > 0$ for all j. Moreover, since $f_{rot_{ij}} = -f_{rot_{ji}}$, if $j \in M_i$ then $i \notin M_j$. Otherwise either $f_{rot_{M_i}} < 0$ or $f_{rot_{M_j}} < 0$ which is impossible.

Define the directed graph $\mathcal{G}_\infty$ with the same nodes as the original graph, but which only includes the edge $i \to j$ is $j \in M_i$. Then, under the first assumption, every edge in $\mathcal{G}_\infty$ is irreversible. That is, if there is an edge from i to j there cannot be any edge back in $\mathcal{G}_\infty$. The second assumption ensures that every node has exactly one edge leaving it in $\mathcal{G}_\infty$. It follows that $\mathcal{G}_\infty$ has as many directed edges as nodes.

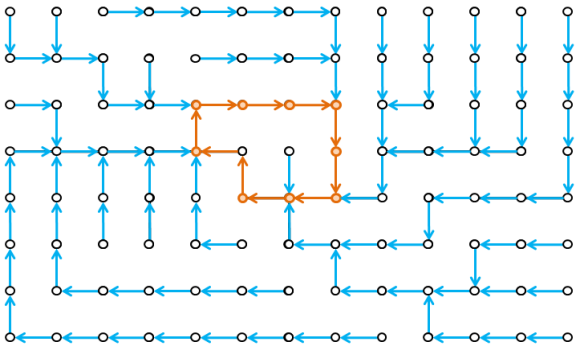

Figure 7.9: The directed graph $\mathcal{G}_\infty$ corresponding to the paths taken by the skeleton process in a strong rotation limit. In this example the original network was a square lattice, the directed graph is connected. The cycle is shown in orange, and the spanning forest in blue.

Consider a connected component of $\mathcal{G}_\infty$. Each connected component has as many edges as it has nodes, so has one more edge than its spanning tree. Since there is at most one directed edge in $\mathcal{G}_\infty$ per undirected edge in $\mathcal{G}$ the edge left out of the spanning tree must be a chord. It follows that every connected component of $\mathcal{G}_\infty$ must contain exactly one cycle. Since every node has an outgoing edge the edges on the cycle must all point along the same direction. Moreover, every other edge in the connected component must be directed so that following the edges in order will lead onto the cycle. An example is illustrated in Figure 7.9

Enumerate the connected components of $\mathcal{G}_\infty$ from 1 to k, and enumerate the corresponding cycles $\mathcal{C}_j$ accordingly. For each cycle let the forest $\mathcal{T}_j$, be the collection of trees, directed towards the cycle $\mathcal{C}_j$, who span the corresponding connected component.

Since every node in $\mathcal{G}_\infty$ has exactly one outgoing edge, we can define a deterministic process which walks in the direction of the outgoing edge leaving each node. If we start at some node i then this process defines a relaxation trajectory $x_j(i)$ where $x_0(i) = i$ and $x_j(i) = M_{x_{j-1}(i)}$. This process is the relaxation process, and the trajectories it produces are relaxation trajectories. A transition along a directed edge in $\mathcal{G}_\infty$ is a relaxation step. Any transition along an edge not in $\mathcal{G}_\infty$ is an activation step, and any consecutive sequence of

activation steps is an activation trajectory. Then, any trajectory X can be separated into a series of relaxation and activation trajectories.

Our objective is to show that, in the large rotation limit, the skeleton process converges to the relaxation process in the sense that the probability of observing more than n consecutive relaxation steps converges to 1 as β goes to infinity for any n, and the probability of observing fewer than m consecutive activation steps converges to one as β goes to infinity for any m. It would follow that, for sufficiently large β, the wait time between consecutive activation trajectories would diverge, and the length of the average activation trajectory trajectory would converge towards one.

To start, rewrite the skeleton transition probabilities:

$$\begin{aligned}\hat{l}_{ij}(\beta) &= \left\{ \begin{array}{l} \dfrac{1}{1+\epsilon_j(\beta)} \text{ if } i = M_j \\ \dfrac{\epsilon_{ij}(\beta)}{1+\epsilon_j(\beta)} \text{ if } i \neq M_j \end{array} \right\} \\ &\simeq \left\{ \begin{array}{l} 1 - \displaystyle\sum_{k \neq M_j} \epsilon_{kj}(\beta) \text{ if } i = M_j \\ \\ \epsilon_{ij}(\beta) \text{ if } i \neq M_j \end{array} \right\} = \left\{ \begin{array}{l} 1 - \mathcal{O}(\exp\left(-\beta \Delta f_j\right)) \text{ if } i = M_j \\ \\ \mathcal{O}(\exp\left(-\beta \Delta f_{ij}\right)) \text{ if } i \neq M_j \end{array} \right\}. \end{aligned} \tag{7.153}$$

Therefore, if the system is in state j the probability of observing a relaxation step is $1 - \mathcal{O}(\exp\left(-\beta \Delta f_j\right))$ and the probability of observing any activation step is $\mathcal{O}(\exp\left(-\beta \Delta f_j\right))$. Therefore the probability of observing a sequence of at least n consecutive relaxation steps starting from node i is:

$$P[n \text{ consecutive relaxation steps} | X_0 = i] \simeq \prod_{j=0}^{n-1} \left(1 - \epsilon_{x_j(i)}(\beta)\right) \tag{7.154}$$

which clearly converges to 1 for any fixed n regardless the initial i.

Next let $\mathcal{A}_n(i)$ be the space of activation trajectories of length n leaving node i. Let

$y = \{y_0 = i, y_1, ...y_{n-1}\}$ be a trajectory. Then the probability of observing any activation trajectory of n or more steps is:

$$\begin{aligned} P[n \text{ consecutive activation steps} | X_0 = i] &\simeq \sum_{y \in \mathcal{A}_n(i)} \prod_{j=0}^{n-1} \epsilon_{y_{j+1}, y_j}(\beta) \\ &= \sum_{y \in \mathcal{A}_n(i)} \exp\left(-\beta \sum_{j=0}^{n-1} \Delta f_{y_{j+1}, y_j}\right) \\ &= \sum_{y \in \mathcal{A}_n(i)} \exp\left(-\beta W[y]\right) \end{aligned} \tag{7.155}$$

where $W[y] = \sum_{j=0}^{n-1} \Delta f_{y_{j+1}, y_j}$ is the work to traverse the activation trajectory. Note that the work is defined relative to Δf rather than f. Also note that since $\Delta f_{y_{j+1}, y_j} > 0$ for all activation steps the work is strictly positive. Therefore the probability of observing any particular activation trajectory is vanishing $\mathcal{O}(\exp\left(-\beta W[y]\right))$.

In a network where each node has finite degree there are always a finite number of activation trajectories of length n leaving any node i. This number is independent of β so the sum must vanish exponentially quickly in β. In the large rotation limit:

$$P[n \text{ consecutive activation steps} | X_0 = i] \simeq \mathcal{O}\left(\exp\left(-\beta \min_{y \in \mathcal{A}_n(i)} \{W[y]\}\right)\right) \to 0 \tag{7.156}$$

so the probability of observing fewer than n consecutive activation steps converges to 1 for any fixed n and i.

It follows that the skeleton process converges to the relaxation process in the large rotation limit. This makes analysis of the steady state of the skeleton process straightforward, provided the $\mathcal{G}_\infty$ is connected.

Assume that $\mathcal{G}_\infty$ is connected. Then $\mathcal{G}_\infty$ contains a unique cycle $\mathcal{C}$ containing $|\mathcal{C}|$ nodes. All relaxation trajectories relax onto the cycle, and once on the cycle never leave. Therefore

the cycle is absorbing in the strong rotation limit so:

$$\lim_{\beta\to\infty} \hat{q}_i(\beta) = \begin{bmatrix} \frac{1}{|\mathcal{C}|} \text{ if } i \in \mathcal{C} \\ 0 \text{ else} \end{bmatrix}. \tag{7.157}$$

Therefore the stationary distribution for the skeleton process converges to a uniform distribution on a cycle $\mathcal{C}$ in the large rotation limit provided: the weights $\rho_{ij}(\beta)$ converge to constants, every node neighbors an edge with nonzero rotational flow, there is a unique edge leaving every node with maximal flow, and the directed graph $\mathcal{G}_\infty$ constructed from these edges is connected. These are all easy conditions to check, and except the first, depend exclusively on the rotational flow.

If $\mathcal{G}_\infty$ is not connected then the limit is singular since, for any finite β the skeleton process admits a unique stationary distribution, but for infinite β the skeleton process is the relaxation process, whose long time average depends on which connected component the process starts in. For large β this implies that each connected component of $\mathcal{G}_\infty$ is nearly absorbing, and the quasisteady state on that component is converging to the uniform distribution on the corresponding cycle. The full steady state converges to an weighted average of the uniform distribution on all the cycles, where the probability of being on any individual cycle is fixed by the mean first passage times between the cycles. This is similar to a network with multiple deep potential wells (cf. [23]). Further discussion of this case is saved for future work.

To find the steady state for the full process, rescale the steady state for the skeleton process by the expected waiting time of each node (recall Equation (7.145)):

$$\lim_{\beta\to\infty} q_i(\beta) = \lim_{\beta\to\infty} \frac{\tau_i(\beta)}{\tau(\beta)} \hat{q}_i(\beta). \tag{7.158}$$

We know that $\lim_{\beta\to\infty} \hat{q}_i(\beta) = 0$ if $i \notin \mathcal{C}$ which would suggest $\lim_{\beta\to\infty} q_i(\beta) = 0$ if $i \notin \mathcal{C}$, however, as β goes to infinity the waiting times all converge to zero. Therefore, if $\tau(\beta)$ converges to zero faster than $\tau_i(\beta)$ the ratio of waiting times may diverge. It is possible, then, that the product $\frac{\tau_i(\beta)}{\tau(\beta)}\hat{q}_i(\beta)$ does not converge to zero, even while $\hat{q}_i(\beta)$ converges to zero. This leads to the possibility that, in the large rotation limit, the support of the steady state of the full process may not be a subset of the support of the steady state distribution of the skeleton process. In fact, depending on the rates at which $\tau_i(\beta)$, $\tau(\beta)$, and $\hat{q}_i(\beta)$ converge to zero it is possible that, in the large rotation limit, the steady state of the full process and the steady state for the skeleton process are disjoint!

The support of the two distributions may differ if, while the process is expected to spend almost all of its steps on $\mathcal{C}$, it is also spend almost all of its time off of $\mathcal{C}$. To see why this is possible consider a network consisting of one cycle and a single node connected to the cycle by a pair of edges. Then it is possible that, in the large rotation limit, the system is expected to make 1000 consecutive complete cycles per fluctuation off the cycle. Then the steady state probability of occupying the single node off the cycle on a given step is on the order of 1/1000th the probability of occupying a node on the cycle. However, if the rate at which the system cycles is also 1000 times faster than the rate at which it leaves the outer node, then the steady state for the full process will be approximately uniform. The process visits the nodes on the cycle 1000 times for every fluctuation of the cycle, but visits each node in the cycle for 1/1000th the time it spends on the fluctuation. Clearly if the rates scale differently it is possible that, in the large rotation limit, the system spends all of its steps on the cycle, but none of its time there.

An example of this kind is presented in [265] in response to numerical results presented by [266, 267]. In [265] example Markov chains are constructed whose skeleton converges to a deterministic process, but where long numerical runs do not capture the steady state

dynamics of the full process. Long numerical runs based on direct simulation [242], or using the first reaction method [243] proceed one transition at a time, so “long" simulation runs are “long" in the sense that many transitions are observed, not necessarily a long time. If the rate of transitions is fast that observing many transitions may not amount to observing a long enough time to observe convergence to steady state.

Therefore, analyzing the strong rotation limit of the steady state distribution of the full process requires good estimators of the asymptotic rates at which $\tau_i(\beta)$, $\tau(\beta)$, and $\hat{q}_i(\beta)$ converge to zero. By comparing the relative rates we can determine whether or not the steady state of the full process reflects the steady state of the skeleton process.

First, separate the expected steady state wait time into a sum over the nodes in the cycle and nodes not in the cycle:

$$\tau(\beta) = \sum_{i \in \mathcal{C}} \tau_i(\beta)\hat{q}_i(\beta) + \sum_{i \notin \mathcal{C}} \tau_i(\beta)\hat{q}_i(\beta)$$

Then:

$$\begin{aligned} q_j(\beta) &= \frac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i \in \mathcal{C}} \tau_i(\beta)\hat{q}_i(\beta) + \sum_{i \notin \mathcal{C}} \tau_i(\beta)\hat{q}_i(\beta)} \\ &= \left[1 + \frac{\sum_{i \in \mathcal{C}, i \neq j} \tau_i(\beta)\hat{q}_i(\beta)}{\tau_j(\beta)\hat{q}_j(\beta)} + \frac{\sum_{i \notin \mathcal{C}, i \neq j} \tau_i(\beta)\hat{q}_i(\beta)}{\tau_j(\beta)\hat{q}_j(\beta)}\right]^{-1} \end{aligned}$$

Then, since both of the sums are strictly positive there are three possibilities:

$$\lim_{\beta \to \infty} q_j(\beta) = \left\{ \begin{array}{l} 0 \text{ if } \dfrac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i \in \mathcal{C}, i \neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to 0 \text{ or } \dfrac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i \notin \mathcal{C}, i \neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to 0 \\ 1 \text{ if } \dfrac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i \in \mathcal{C}, i \neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to \infty \text{ and } \dfrac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i \notin \mathcal{C}, i \neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to \infty \\ \text{some number between 0 and 1 otherwise} \end{array} \right\}. \tag{7.159}$$

Let's find sufficient conditions to ensure that:

$$\lim_{\beta\to\infty} \hat{q}_i(\beta) = 0 \tag{7.160}$$

for all $i \notin \mathcal{C}$. Define the numerical support of a distribution:

$$\text{supp}_\epsilon(p) = \{x | p(x) >= \epsilon\}. \tag{7.161}$$

Then we are looking for a sufficient condition to guarantee:

$$\lim_{\beta\to\infty} \text{supp}_\epsilon(q(\beta)) \subseteq \lim_{\beta\to\infty} \text{supp}_\epsilon(\hat{q}(\beta))$$

for any $\epsilon > 0$. If the system satisfies this condition then it is rare-fluctuation stable, since in the limit rare fluctuations in the skeleton process do not contribute undue weight to the steady state of the full process. If the support of the full process contains the support of the skeleton the the process is rare-fluctuation neutral since rare fluctuations contribute as much to the steady state as the relaxation process does. If the support of the full process is disjoint from the support of the skeleton then the process is rare-fluctuation unstable since rare fluctuations contribute the majority of the steady state.

A system is rare-fluctuation stable if and only if $\lim_{\beta\to\infty} \hat{q}_j(\beta) = 0$ for all j outside of $\mathcal{C}$. This requires that, for all j outside of $\mathcal{C}$ either $\frac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i\in\mathcal{C}, i\neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to 0$ or $\frac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i\notin\mathcal{C}, i\neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to 0$. Therefore it would be sufficient if $\frac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i\in\mathcal{C}, i\neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to 0$ for all j outside of the cycle. This is also necessary, since there must be some set of j outside the cycle such that $\tau_j(\beta)\hat{q}_j(\beta)$ goes to zero slower than, or as slow, as on all other nodes outside the cycle, in which case it is impossible that $\frac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i\notin\mathcal{C}, i\neq j} \tau_i(\beta)\hat{q}_i(\beta)} \to 0$ for all $j \notin \mathcal{C}$. Therefore, a system

is rare-fluctuation stable if and only if, for all $j \notin \mathcal{C}$:

$$\frac{\tau_j(\beta)\hat{q}_j(\beta)}{\sum_{i\in\mathcal{C}}\tau_i(\beta)\hat{q}_i(\beta)} \to 0. \tag{7.162}$$

This is a natural condition. A system is rare fluctuation stable if the expected time spent on any fluctuation off the cycle can be made arbitrarily small relative to the expected time spent on the cycle by picking β sufficiently large. Now, for all i on the cycle $\hat{q}_i(\beta) \to 1/|\mathcal{C}|$. Therefore $1/\hat{q}_i(\beta)$ converges to $|C|$. Then, the limit of the ratio converges to zero if and only if $|\mathcal{C}|\frac{\tau_j(\beta)}{\sum_{i\in\mathcal{C}}\tau_i(\beta)}\hat{q}_j(\beta) \to 0$. The convergence to zero is clearly independent of $|\mathcal{C}|$ so rare-fluctuation stability requires:

$$\frac{\tau_j(\beta)}{\sum_{i\in\mathcal{C}}\tau_i(\beta)}\hat{q}_j(\beta) \to 0. \tag{7.163}$$

In the large β limit:

$$\begin{aligned}
&\lim_{\beta\to\infty}\tau_j(\beta) = \mathcal{O}(\exp\left(-\beta f_{rot_{M_j}}\right)) \\
&\lim_{\beta\to\infty}\sum_{i\in\mathcal{C}}\tau_i(\beta) = \mathcal{O}(\exp\left(-\beta\min_{i\in\mathcal{C}}\{f_{rot_{M_i}}\}\right)
\end{aligned} \tag{7.164}$$

Therefore:

$$\frac{\tau_j(\beta)}{\sum_{i\in\mathcal{C}}\tau_i(\beta)}\hat{q}_j(\beta) = \mathcal{O}\left(\exp\left(\beta\left(\min_{i\in\mathcal{C}}\{f_{rot_{M_i}}\} - f_{rot_{M_j}}\right)\right)\right). \tag{7.165}$$

If $f_{rot_{M_j}} \geq \min_{i\in\mathcal{C}}\{f_{rot_{M_i}}\}$ then the ratio of the waiting times does not diverge, so $q_i(\beta)$ converges to zero. However, if $f_{rot_{M_j}} < \min_{i\in\mathcal{C}}\{f_{rot_{M_i}}\}$ then the ratio of the waiting times diverges, so $q_i(\beta)$ goes to zero if and only if:

$$\lim_{\beta\to\infty} \exp\left(\beta\left(\min_{i\in\mathcal{C}}\{f_{rot_{M_i}}\} - f_{rot_{M_j}}\right)\right)\hat{q}_j(\beta) = 0. \tag{7.166}$$

This gives a necessary and sufficient condition on the asymptotic rate at which $\hat{q}_j(\beta)$ must converge to zero in order to guarantee that the full process is rare-fluctuation stable. Therefore, in order to check if a system is rare-fluctuation stable we will have to develop bounds on the asymptotic rate at which the steady state probability at all nodes off the cycle converges to zero.

For any node $i \notin \mathcal{C}$ define the asymptotic rate of convergence to zero to be the limit, $-\lim_{\beta\to\infty} \frac{1}{\beta}\log\left(\hat{q}_i(\beta)\right)$. To work out these asymptotic rates consider the steady state equation for a discrete time process:

$$\hat{L}(\beta)\hat{q}(\beta) = \hat{q}(\beta). \tag{7.167}$$

It follows that, for any n:

$$\hat{L}(\beta)^n\hat{q}(\beta) = \hat{q}(\beta). \tag{7.168}$$

The product $\hat{L}^n\hat{q}$ can be rewritten in terms of all paths length n. Let $Y_{ij}(n)$ be the space of all paths from j to i of length n. Define the measure on paths:

$$\pi(y|\beta) = \prod_{j=1}^{|y|} \hat{l}_{y_{j+1},y_j}(\beta). \tag{7.169}$$

Then:

$$\hat{q}_i(\beta) = \sum_j \left[\sum_{y\in Y_{ij}(n)} \pi(y|\beta)\right]\hat{q}_j(\beta) \tag{7.170}$$

That is, the steady state probability at any node is given by the sum over the probability

of reaching that node from any other node, which can be broken into a sum over all paths between pairs of nodes. In the large rotation limit the probability of any path converges to the exponential of the work to traverse the path where work is evaluated against Δf:

$$\pi(y|\beta) \simeq \left[\prod_{j=1}^{|y|} \frac{\rho_{y_{j+1},y_j}}{\rho_{M_{y_j}}}\right] \exp\left(-\beta W[y]\right). \tag{7.171}$$

Fix, i and j and n. Then let:

$$\{y_{ij}(n)\} = \text{argmin}_{y \in Y_{ij}(n)}\{W[y]\} \tag{7.172}$$

be the set of the most likely paths from j to i in n steps. Note that there may be more than one path in this set, but the number of such paths is finite since n is finite. Also let:

$$\Delta W[y]_{ij}(n) = W[y] - W[y_{ij}(n)] > 0 \tag{7.173}$$

for some optimal path in the set.

Then, in the large β limit:

$$\pi(y|\beta) \simeq \exp\left(-\beta W[y_{ij}(n)]\right) \left\{ \begin{array}{l} \displaystyle\prod_{j=1}^{|y|} \frac{\rho_{y_{j+1},y_j}}{\rho_{M_{y_j}}} \text{ if } y \in \{y_{ij}(n)\} \\ \displaystyle\left[\prod_{j=1}^{|y|} \frac{\rho_{y_{j+1},y_j}}{\rho_{M_{y_j}}}\right] \exp\left(-\beta \Delta W[y]_{ij}(n)\right) \text{ if } y \notin \{y_{ij}(n)\} \end{array} \right\} \tag{7.174}$$

Since $\Delta W[y]_{ij}(n) > 0$ the distribution of paths from j to i in n steps collapses onto the

most likely paths from j to i in n steps. Therefore:

$$\left[\sum_{y\in Y_{ij}(n)}\pi(y|\beta)\right] \simeq \left[\sum_{y\in\{y_{ij}(n)\}}\prod_{j=1}^{|y|}\frac{\rho_{y_{j+1},y_j}}{\rho_{M_{y_j}}}\right]\exp\left(-\beta W[y_{ij}(n)]\right) \tag{7.175}$$

For any i and n let $W_{i\mathcal{C}}(n)$ be the minimum work needed to reach i from $\mathcal{C}$ in n steps. Now consider some path y from $j\in\mathcal{C}$ to $i\notin\mathcal{C}$. Let $j-1$ be the previous node in the cycle. Then the path from $j-1$ to j to i along y takes the same amount of work as y itself since it takes no work to traverse the loop. Therefore, for any path from the cycle to a node not in the cycle, all paths that consist of cycling the loop for an arbitrary number of steps then following a specified path take the same amount of work, so have probabilities that decay at the same asymptotic rate. It follows that $W_{i\mathcal{C}}(n+1)\leq W_{i\mathcal{C}}(n)$. Therefore the minimum work from $\mathcal{C}$ to i is nonincreasing as we increase the length of the paths. If the network is finite then this sequence must have an nonzero minimum that is first achieved for some finite n since any path from $\mathcal{C}$ to i that uses more than $V-|\mathcal{C}|$ steps must be equivalent to a shorter path plus a cycle. Since the work is additive and the work to traverse any edge is nonnegative, adding a cycle to a path never decreases the work to traverse the path. Therefore, the minimum work to traverse the optimal path from $\mathcal{C}$ to i of length longer than $V-|\mathcal{C}|$ must be greater than or equal to the minimum work to traverse the optimal path from $\mathcal{C}$ to i of length less than $V-|\mathcal{C}|$. Let $W_{i\mathcal{C}}$ be the minimum work to go from $\mathcal{C}$ to i. Then, if we pick $n>V$ we can guarantee that a path of this work from some j in $\mathcal{C}$ to i is included in the set of paths:

Then, since $\hat{q}_j(\beta)\to\frac{1}{|\mathcal{C}|}$ in the large rotation limit:

$$\hat{q}_i(\beta)\simeq\left[\sum_{j\notin\mathcal{C}}\omega_{ij}(n)\exp\left(-\beta W[y_{ij}(n)]\right)\hat{q}_j(\beta)\right]+\frac{1}{|\mathcal{C}|}\exp\left(-\beta W_{i\mathcal{C}}\right) \tag{7.176}$$

where $\omega_{ij}(n)$ is the ratio of the conductances along any given path to the optimal path. Note that we do not sum over all $j \in \mathcal{C}$ because for any n there is a unique $j \in \mathcal{C}$ that is precisely n steps away on the optimal path.

Notice that the second half of the equation is now independent of n. The right hand side, however, depends on n. If n is sufficiently large then it is possible to reach i from j by following a relaxation path from j to $\mathcal{C}$ then by following the optimal path from $\mathcal{C}$ to i. This path also has work $W_{i\mathcal{C}}$ since the relaxation steps come for free. This must be the least work it takes to go from j off the cycle to the cycle, then to i, since any other path either takes more work to reach the cycle, or more work to reach i from the cycle. For small n it is possible that there is a path from j to i without reaching the cycle first that takes less work than $W_{i\mathcal{C}}$. However, if n is sufficiently large (larger than V) then any path from j to i without relaxing to $\mathcal{C}$ first, must include a cycle that is not $\mathcal{C}$. Since $\mathcal{C}$ is the only cycle in $\mathcal{G}_\infty$, it must take positive work to traverse any such cycle. There is some combination of a path from j to i and cycle that leaves and returns to the path, both without reaching $\mathcal{C}$, that minimizes the work for any given n. However, if we increase n we need to either take a different path, different cycle, or repeat the same cycle multiple times. Therefore the minimum work from j to i is greater than or equal to the minimum work to go from j to i without reaching $\mathcal{C}$ plus the smallest work needed to traverse any cycle that is not $\mathcal{C}$ times n. It follows that the work to move from j to i without reaching $\mathcal{C}$ first must diverge as n goes to infinity. Then there is necessarily some n large enough that the cheapest path from j to i is the path that first relaxes to $\mathcal{C}$. Then:

$$\hat{q}_i(\beta) \simeq \left[\sum_{j \notin \mathcal{C}} \exp\left(-\beta W_{i\mathcal{C}}\right) \hat{q}_j(\beta)\right] + \frac{1}{|\mathcal{C}|} \exp\left(-\beta W_{i\mathcal{C}}\right)$$

Then, since $\hat{q}_j(\beta)$ converges to zero as β goes to infinity:

$$\hat{q}_i(\beta) \simeq \frac{1}{|\mathcal{C}|} \exp\left(-\beta W_{i\mathcal{C}}\right) \tag{7.177}$$

Therefore the steady state probability of occupying any node that is not on $\mathcal{C}$ decays to zero at rate equal to the minimum work to go from the cycle to i:

$$-\lim_{\beta\to\infty} \frac{1}{\beta} \log\left(\hat{q}_i(\beta)\right) = \min_{y\in Y_{i\mathcal{C}}}\{W[y]\} = W_{i\mathcal{C}}. \tag{7.178}$$

Lemma 36 (Rare fluctuation stability)**.** *A purely rotational network whose geometric mean transition rates, $\rho_{ij}(\beta)$, converge to constants in β, satisfying the three assumptions:*

- Assumption 1: *For all i there exists a j such that $f_{rot_{ij}} \neq 0$,*
- Assumption 2: *For each i there is a unique j that maximizes $f_{rot_{ij}}$,*
- Assumption 3: *$\mathcal{G}_\infty$ is connected,*

is rare-fluctuation stable:

$$\lim_{\beta\to\infty} \textit{supp}_\epsilon(q(\beta)) \subseteq \lim_{\beta\to\infty} \textit{supp}_\epsilon(\hat{q}(\beta)) \tag{7.179}$$

if and only if the minimal work to move from $\mathcal{C}$ to $i \notin \mathcal{C}$ is greater than the difference between the largest flow leaving the slowest node on the cycle minus the largest flow leaving i:

$$\min_{y\in Y_{i\mathcal{C}}}\{W[y]\} > \min_{j\in\mathcal{C}}\{f_{M_j}\} - f_{M_i}. \tag{7.180}$$

for all $i \notin \mathcal{C}$.

Note that, beyond the first condition, every other assumption and requirement depends purely on f_{rot}. In order to check these requirements follow the following steps:

1. Check that the conductances $\rho_{ij}(\beta)$ do not diverge in the strong rotation limit
2. For each node find $f_{M_i} = \max_{j \in \mathcal{N}_i}\{f_{rot_{ij}}\}$ and $M_i = \text{argmax}_{j \in \mathcal{N}_i}\{f_{rot_{ij}}\}$
3. Check whether $f_{M_i} = 0$ for any i and check whether $|M_i| = 1$ for all i
4. Given M_i, form the directed network $\mathcal{G}_\infty$ whose edges correspond to the edges with max flow leaving each node
5. Determine whether or not $\mathcal{G}_\infty$ is connected and find the cycle $\mathcal{C}$ (this can be accomplished by searching downhill from a randomly drawn initial node until completing a cycle, then using a purely uphill breadth first search starting from the cycle. If this does not reach every node in the network then the directed network is not connected.)
6. Find $\min_{j \in \mathcal{C}}\{f_{M_j}\}$
7. Define $\Delta f_{ij} = f_{M_j} - f_{rot_{ij}}$ and the work over paths $W[y] = \sum_{j=1}^{|y|} \Delta f_{ij}$
8. Use an optimal tree search (see Appendix D) to find the minimum work to reach all $i \notin \mathcal{C}$ from some $j \in \mathcal{C}$.
9. Compare $\min_{y \in Y_{i\mathcal{C}}}\{W[y]\}$ to $\min_{j \in \mathcal{C}}\{f_{M_j}\} - f_{M_i}$ for all $i \notin \mathcal{C}$
10. If $\min_{y \in Y_{i\mathcal{C}}}\{W[y]\} > \min_{j \in \mathcal{C}}\{f_{M_j}\} - f_{M_i}$ for all $i \notin \mathcal{C}$ then the network is rare-fluctuation stable.

If the network is rare-fluctuation stable then it is easy to approximate $q(\beta)$ for large β. For all $i \notin \mathcal{C}$, the steady state $q_i(\beta)$ converges to zero exponentially fast in β. Therefore we

approximate $q_i(\beta)$ with zero for all $i \notin \mathcal{C}$. For $i \in \mathcal{C}$:

$$q_i(\beta) \simeq \frac{\tau_i(\beta)}{\sum_{j\in\mathcal{C}} \tau_j(\beta)} \simeq \frac{\rho_{M_i}^{-1} \exp\left(-\beta f_{M_i}\right)}{\max_{j\in\mathcal{C}}\{\rho_{M_j}^{-1} \exp\left(-\beta f_{M_j}\right)\}}. \tag{7.181}$$

Let $m_{\mathcal{C}}$ be the set of $j \in \mathcal{C}$ that minimizes f_{M_j}. Then:

$$q_i(\beta) \simeq \frac{r_{M_i}}{R_{m_{\mathcal{C}}}} \exp\left(-\beta(f_{M_i} - f_{m_{\mathcal{C}}})\right). \tag{7.182}$$

where $r = 1/\rho$ are the per capita resistances, $R_{m_{\mathcal{C}}} = \sum_{j\in m_{\mathcal{C}}} r_{M_j}$ is the total resistance, and $f_{m_{\mathcal{C}}} = \min_{j\in\mathcal{C}}\{f_{M_j}\}$.

Therefore, if the network is rare-fluctuation stable, then the steady state of the full process converges to zero everywhere outside $\mathcal{C}$, and on all nodes on $\mathcal{C}$ that do not precede the slowest edge (or edges) on $\mathcal{C}$. The probability on these nodes decays exponentially at rate fixed by the difference between the flow leaving the node of interest and the weakest flow on the cycle. If there is a unique edge on the cycle with the smallest outgoing flow, then the steady state converges to a delta distribution at the preceding node, with exponentially decaying probability on all other nodes. This is the general large rotation limit when $\mathcal{G}_\infty$ is connected since any other limit requires exact equalities between forces on edges. If any two edges on the cycle have the same outgoing force then the steady state converges to the distribution of weights on the nodes preceding these edges.

That said, since f_{rot} is lower dimensional than the space of edges, there are a variety of interesting cases when these equalities play an important role. As an extreme example, if the network is a single loop then f_{rot} is necessarily the same on all edges. It follows that, in the strong rotation limit the steady state distribution converges to the ratio of the inverse

conductances:

$$q_i(\beta) \simeq \frac{r_i}{R}, \quad R = \sum_j r_j, \quad r_j = \frac{1}{\rho_j} \tag{7.183}$$

as we had discovered when studying networks with isolated loops (see Section 7.2.2).

To estimate the steady state for networks that are not rare-fluctuation stable but obey all the other conditions we note that the steady state distribution is asymptotically proportional to:

$$q_i(\beta) \propto \frac{1}{\rho_{M_i}} \exp\left(\beta\left(\min_{j\in\mathcal{C}}\{f_{M_j}\} - f_{M_i} - \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right)\right). \tag{7.184}$$

Now, the support of the steady state converges to:

$$\lim_{\beta\to\infty} \text{supp}_\epsilon\{q(\beta)\} \subseteq \text{argmax}_i\{\min_{j\in\mathcal{C}}\{f_{M_j}\} - f_{M_i} - \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\} \tag{7.185}$$

with equality achieved for sufficiently small $\epsilon > 0$. On this set the steady state converges to the ratio of the resistances:

$$\lim_{\beta\to\infty} q_i(\beta) \propto r_i. \tag{7.186}$$

Off this set the steady state probability converges exponentially to zero:

$$q_i(\beta) = \mathcal{O}\left(-\beta\left(\left[f_{M_i} + \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right] - \text{argmin}_k\{f_{M_k} + \min_{y\in Y_{k\mathcal{C}}}\{W[y]\}\}\right)\right).$$

Therefore, for large β and i off the asymptotic support:

$$\lim_{\beta\to\infty} \frac{1}{\beta}\log\left(q_i(\beta)\right) = -\left(\left[f_{M_i} + \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right] - \text{argmin}_k\{f_{M_k} + \min_{y\in Y_{k\mathcal{C}}}\{W[y]\}\}\right). \tag{7.187}$$

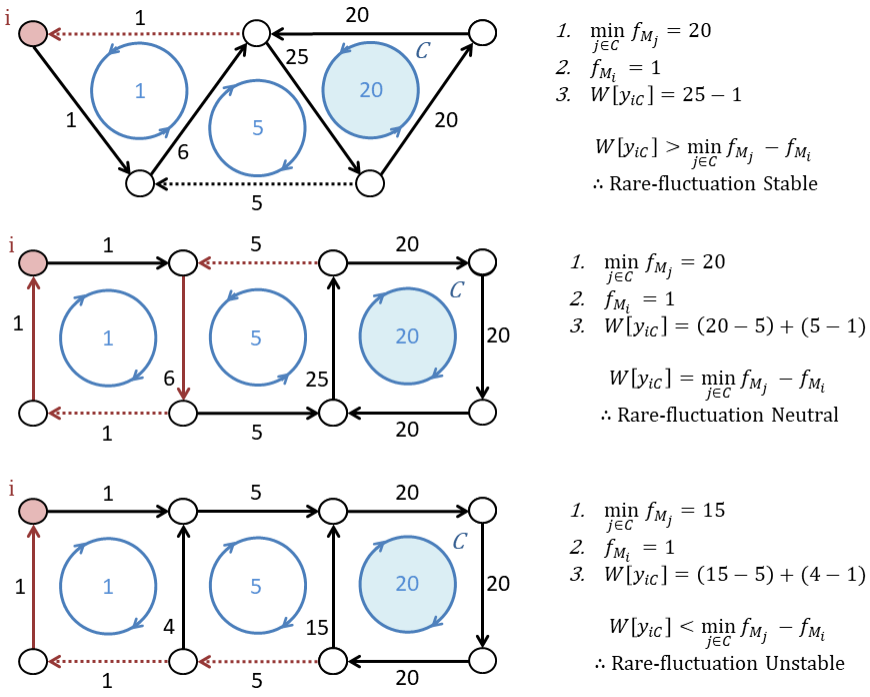

Figure 7.10: Rare-fluctuation stability of three examples. The first is stable, the second is neutral, and the third is unstable.The solid arrow represent $\mathcal{G}_\infty$, the dashed arrows represent activation steps. The red shaded node is node i and the blue shaded circle is $\mathcal{C}$ the red edges represent the optimal activation trajectory.

So, for any network satisfying the first constraint (geometric mean transition rates converge to finite nonzero constants), and the three assumptions, we can exactly solve for the large rotation limit of the steady state of the full process and skeleton process, and solve for the asymptotic rate at which the steady state converges to zero off the support for both processes. This is a fairly strong characterization, and can be accomplished purely by studying the rotational forces. In particular, it requires finding the directed graph $\mathcal{G}_\infty$ and the optimal spanning tree that minimizes the work to move from the cycle to any node off the cycle. This can be accomplished using two straightforward search routines.

Three examples are shown in Figure 7.10 corresponding to a rare-fluctuation stable, rare-fluctuation neutral, and rare-fluctuation unstable processes.

Note that if node i can be reached from node j by a relaxation trajectory then the minimal work to reach i is always less than or equal to the minimal work to reach j. Therefore, the optimal spanning tree will typically consist of a series of activation trajectories that reach a subset of nodes, followed by a series of branches formed by the relaxation trajectories leaving the activation trajectories and terminating before they arrive at an alternative activation trajectory. The work to reach any node on these branches is constant, so the asymptotic rate at which the skeleton steady state converges to zero on these branches is constant, and in the full process is proportional to the time spent on each. This is analogous to the circulant left over by the quasipotential (see [23]), since it defines a set of level sets (or isoclines) in the work to escape an attracting set. These level sets all share the same work to reach, and the asymptotic probability of occupying any node in a level set is equal to any other node in the level set (for the skeleton process). Moreover the rate at which the skeleton steady state converges to zero off the asymptotic support is exponential in the work on the optimal tree - exactly the behavior of the steady state of an SDE in the small noise limit relative to the quasipotential.

We can relax the original constraint that the geometric transition rates converge to nonzero finite constants. If instead we constrain the weights so that the expected time to remain in any node converges to a finite nonzero constant then, in the limit:

$$\lim_{\beta\to\infty} \tau_j(\beta) = \tau_j > 0 \tag{7.188}$$

so the steady state of the full process converges to:

$$\lim_{\beta\to\infty} q_i(\beta) = \frac{\tau_i \hat{q}_i(\beta)}{\sum_j \tau_j \hat{q}_j(\beta)} \simeq \frac{\tau_i}{\tau_{\mathcal{C}}} \hat{q}_i \tag{7.189}$$

where: $\tau_{\mathcal{C}} = \frac{1}{|\mathcal{C}|} \sum_{j\in\mathcal{C}} \tau_j$. Therefore, under the third constraint on the conductances (finite

nonzero waiting times in each node), the steady state for the full process is just the steady state for the skeleton process scaled by the expected waiting times. Since these remain finite the full process is automatically rare-fluctuation stable.

What about the other assumptions?

The third assumption on f_{rot} was that the directed graph $\mathcal{G}_\infty$ formed by the set of edges with maximal flow is connected. If $\mathcal{G}_\infty$ is not connected then the steady state of the skeleton process is singular in the limit. That is, a unique steady state exists for all finite β, but does not exist for infinite β. This is because the process converges to a deterministic process where the long term behavior of the process depends on the initial condition. For finite β the actual steady state depends on the rate of rare transitions from one cycle to another. These depend on the probabilities of rare-fluctuations, so, like the probabilities of nodes off the cycle, are controlled by the work along the most likely path from one cycle to the basin of attraction of another. We can use these probabilities to work out the expected rate of transition from any cycle to any other cycle. In the large β limit we are expected to spend arbitrarily long within each basin, so we can coarse grain the Markov process on the original state space to a discrete time Markov process on the space of cycles, with transition probabilities which scale exponentially in the minimum work to move from one cycle to another. These probabilities converge to zero exponentially fast in β. In general each transition between connected components occurs at a different time scale, so timescale separation arguments could be used (see [268]) to estimate the steady state.

What about the second assumption on f_{rot}? This assumption was essential since it allowed us to replace the skeleton process with the relaxation process (plus very occasional fluctuations) in the limit. If this assumption is not enforced then the skeleton process does not converge to a deterministic process, so the analysis becomes more difficult. Then the analysis based on the difference in spanning trees presented in Section 7.3.3 should be used.

What about the second constraint on the limiting behavior of the conductances? The second constraint required that the arithmetic mean transition of each pair of forward and backward rates converged to a finite nonzero value. This constraint makes the analysis considerably more difficult because the skeleton process does not converge to a deterministic relaxation process. Consider a pair of forward and backward rates l_{ij}, l_{ji}. In order for the ratio of the two to diverge in the large rotation limit, but the mean to remain constant, they must converge to $l_+, 0$ where l_+ is the arithmetic mean of the transition rates. Therefore, the process becomes irreversible in the limit, but the skeleton process does not become any closer to deterministic. In fact, this limit moves to edge of the space of Markov chains that can be described by the potential framework since it converges to an irreversible process with finite mean transition rates. These rates are independent of the size of the edge flow and only depends on the sign of the forces and the constraint on the arithmetic mean. Therefore the potential framework does not apply naturally to this limit.

The Near Deterministic Limit and the Network Quasipotential

We have shown that the steady state for an arbitrary Markov process with reversible transitions can be computed by first scaling by the equilibrium distribution then solving for the steady state to a purely rotational process. We then developed asymptotic methods for approximating the steady state for a purely rotational process when rotation is weak and strong. When it is strong we showed that assumptions about the asymptotic behavior of the conductances led to different steady states. In particular, if the conductances are assumed to scale so that the expected waiting time in each node converges to a finite nonzero constant then the steady state can be effectively approximated from the steady state for the skeleton process. The steady state for the skeleton process was easy to approximate under some basic assumptions on the structure of the rotational forces. We showed that if

it there is a unique edge leaving each node with maximal rotational force then the steady state could be approximated by computing the probability of rare fluctuations away from a deterministic relaxation process. We then saw that these asymptotic probabilities obeyed the same properties as the asymptotic quasisteady state near a stable attractor in an SDE, and saw that the network quasipotential played the same role as the quasipotential (see [23]). The quasipotential applies in a small noise limit, whereas the network quasipotential applied in a joint strong rotation, fixed waiting time limit. Both cases are near-deterministic limits with expected waiting times that do not converge to zero in the limit. Here we would like to generalize this approach to strongly forced networks whose waiting times do not diverge.

When we originally considered the strongly forced limit (Section 7.3.3) we assumed the conductances, the geometric mean of each forward and backward transition rate, were fixed. In practice this is a strange limit, since the forward rates diverge to infinity. Then the rate at which the system moves becomes infinitely fast. As a result, fluctuations away from the deterministic relaxation process can take arbitrarily longer than relaxation trajectories, so the steady state does not necessarily reflect the underlying deterministic nature of the skeleton process. We provided an example of a rare-fluctuation unstable network where the support of the steady state of the full process was disjoint from the support of the steady state of the skeleton process in the limit. This difficulty made computing the steady state in the high temperature limit different from computing the steady state in a small noise limit. In a small noise limit the rate at which the system moves does not diverge, since a deterministic ODE has a characteristic rate of evolution. Therefore it is not surprising that the strong forcing limit considered in Section 7.3.3 did was not governed by a quasipotential of the same form. Here we will show that the appropriate analogy to small noise is a strongly forced, nonzero waiting time limit, which we will call the "near

deterministic" limit, since in the limit the skeleton process becomes close to deterministic and the corresponding deterministic relaxation process gives good approximation to the steady state of the full process. We will then show that in this limit the full steady state is given by a network quasipotential which closely matches the quasipotential used in the continuum.

In the near deterministic limit the transition rates take the form:

$$l_{ij}(\beta) = \rho_{ij}(\beta) \exp(\beta f_{ij})$$

where $\rho_{ij}(\beta) = \rho_{ji}(\beta)$ and are chosen so that:

$$0 < \lim_{\beta \to \infty} \sum_{i \in \mathcal{N}_j} \rho_{ij}(\beta) \exp(\beta f_{ij}) = \tau_j < \infty \tag{7.190}$$

for any pair of edges leaving j such that $f_{ij} > f_{kj}$:

$$\lim_{\beta \to \infty} \frac{l_{kj}(\beta)}{l_{ij}(\beta)} = \mathcal{O}(\exp(-\beta(f_{ij} - f_{kj}))). \tag{7.191}$$

and it is assumed that for all nodes neighboring edges with nonzero edge flows there is a unique edge M_i which maximizes the outgoing forces. Note that we require that the waiting time is nonzero, but allow it to be arbitrarily long. In a nearly deterministic process we primarily care about the rate of forward reactions since almost all reactions are close to irreversible. We would like to avoid the situation in which forward reactions occur infinitely fast, or waiting times converge to zero, since in that situation rare fluctuations can carry more weight in the limit than the expected trajectory. Moreover, in a small noise limit it is entirely possible that the process waits for a long time somewhere (say near a stable state), however the process should not move infinitely fast when noise is removed.

Then the near deterministic limit is a limit in which all backward rates vanish, all forward rates converge to finite constants, and all but one forward rate leaving each node converge to zero.

The near deterministic limit requires that the conductances scale according to:

$$\rho_{i,j} = \mathcal{O}(\exp\left(-\beta \max\{f_{M_j,j}, 0\}\right)) \tag{7.192}$$

to maintain finite waiting times. Let:

$$\lim_{\beta\to\infty} \rho_{i,j}(\beta) = \frac{1}{\tau_j} \exp\left(-\beta \max\{f_{M_j,j}, 0\}\right) \tag{7.193}$$

where τ_j is the expected waiting time in node j in the limit.

Under these assumptions the waiting time to stay at node j converges to τ_j if there is a forward reaction leaving j and zero otherwise. The transition probabilities for the skeleton process converge to:

$$\hat{l}_{ij}(\beta) \simeq \begin{Bmatrix} 1 - \epsilon_j(\beta) \text{ if } i = M_j \\ \epsilon_{ij}(\beta) \text{ if } i \neq M_j \end{Bmatrix}$$

where:

$$\epsilon_j(\beta) = \sum_{i\neq M_j} \epsilon_{ij}(\beta), \quad \epsilon_{ij}(\beta) = \mathcal{O}(-\beta(f_{M_j} - f_{ij})).$$

Then, under the assumption that $|M_j| = 1$ for all nodes we can define the directed graph $\mathcal{G}_\infty$ which has an edge leaving every node neighboring an edge with a nonzero force. We assume that all nodes neighbor an edge with nonzero force, so that $\mathcal{G}_\infty$ has one edge leaving every node. Since the forces need not be purely rotational it is no longer true that the directed graph contains no reversible transitions. If we extend the definition of a cycle to include any sequence of directed edges that start and return to the same node then the

graph still breaks into a set of connected components each consisting of a series of directed trees that relax onto a cycle. The only difference now is that the cycle might be a 2-cycle (alternation back and forth between a pair of nodes). Since the waiting times all converge to finite constants the mixing time to a unique steady state will diverge if there are multiple disjoint components. In that case the steady state is singular in the deterministic limit, so each basin of attraction should be studied independently to find the quasisteady state distributions.

Focus on a particular basin of attraction. Let $W[y]$ be the work to traverse a path y in the skeleton process (against Δf as defined in Equation (7.147) instead of f). Let $\mathcal{C}$ be the cycle, and consider the steady state for the skeleton process. Then, following the same logic as before, for $i \in \mathcal{C}$:

$$\lim_{\beta\to\infty} \hat{q}_i(\beta) = \frac{1}{|\mathcal{C}|} \tag{7.194}$$

and for $i \notin \mathcal{C}$:

$$\lim_{\beta\to\infty} \frac{1}{\beta} \log\left(\hat{q}_i(\beta)\right) = -\min_{y\in Y_{i\mathcal{C}}}\{W[y]\} \tag{7.195}$$

where $Y_{i\mathcal{C}}$ is the space of all paths from the cycle to the node i not containing any loops. Then:

$$\lim_{\beta\to\infty} q_i(\beta) \propto \tau_i(\beta) \exp\left(-\beta \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right) = \exp\left(\beta\left(\frac{1}{\beta}\log(\tau_i(\beta)) - \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right)\right) \tag{7.196}$$

for $i \notin \mathcal{C}$ and is proportional to the distribution of waiting times τ_i on the loop. This is important if there is a node on the cycle where all the incoming forces are positive, since on this node the expected waiting time is allowed to diverge. In particular, if $\max\{f_{M_j}, 0\} = 0$ then $\frac{1}{\beta}\log(\tau_j(\beta)) \to f_{M_j}$. Otherwise $\tau_j(\beta)$ converges so $\frac{1}{\beta}\log(\tau_j(\beta)) \to 0$. Therefore, in

the limit:

$$\lim_{\beta\to\infty} q_i(\beta) \propto \tau_i(\beta)\exp\left(-\beta \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right) \simeq \exp\left(\beta\left(\max\{f_{M_j},0\} - \min_{y\in Y_{i\mathcal{C}}}\{W[y]\}\right)\right) \tag{7.197}$$

Suppose the basin of attraction converges to a two-cycle and there is one node in the two cycle that has no positive flow leaving it. This is equivalent to a stable node in an ODE. Label this node 1. Let f_{M_1} be the negative flow pushing against the fluctuation from 1 to M_1. Then $\tau_1(\beta)$ diverges proportional to $\exp\left(-\beta f_{M_1}\right)$ [5]. This is the only waiting time that diverges exponentially since all other nodes have an edge leaving them with positive flow. Then $q_i(\beta)/q_1(\beta)$ is, at best, proportional to $\frac{\tau_{M_1}(\beta)}{\tau_1(\beta)}\hat{q}_i(\beta) \propto \exp\left(-\beta f_{1,M_1}\right)\hat{q}_i(\beta)$. Therefore the steady state of the full process is proportional to:

$$\lim_{\beta\to\infty}\frac{q_i(\beta)}{q_1(\beta)} \propto \begin{bmatrix} 1 \text{ if } i=1 \\ \exp\left(-\beta f_{1,M_1}\right) \text{ if } i = M_1 \\ \exp\left(-\beta(f_{1,M_1} + \min_{y\in Y_{i\mathcal{C}}}\{W[y]\})\right) \end{bmatrix}. \tag{7.198}$$

Let $S(i)$ be the function:

$$S(i) = \begin{bmatrix} -f_{1,M_1} \text{ if } i=1 \\ 0 \text{ if } i = M_1 \\ \min_{y\in Y_{i\mathcal{C}}}\{W[y]\} \end{bmatrix}. \tag{7.199}$$

Then S plays the role of the quasipotential for the system. It is the minimal work to move to any node i from node 1. The work to move anywhere other than 1 is at least f_{1,M_1}, and work is evaluated against the $\Delta f_{ij} = f_{ij} - \max\{f_{M_j},0\}$. Then the effective potential

[5] remember $f_{M_1} < 0$

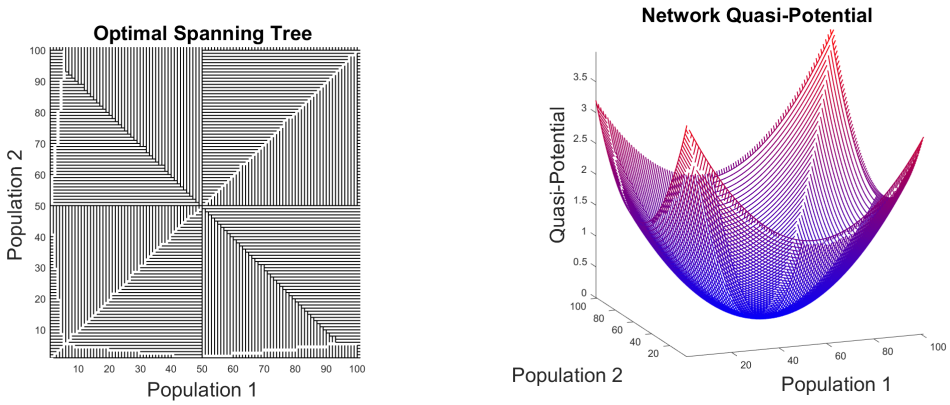

Figure 7.11: An optimal spanning tree and associated quasipotential for a two dimensional lattice with $x_0 = K = [50, 50]$ and with edge rates set to approximate an OU process with constant noise variance $\sigma = 8$, and with a symmetric drift matrix.

(negative log of the steady state) converges to the network quasipotential:

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log(q_i(\beta)) = -S_i + \min_i S_i. \tag{7.200}$$

The network quasipotential can be found by applying the optimal spanning tree algorithm developed in Appendix D. An example is illustrated in Figure 7.11

Therefore, by considering a limit in which the probability of backward transitions vanishes, forward transitions occur at finite (possibly very slow) rates, and every node has a unique neighbor that the process is most likely to move to, we arrive at a limit in which a discrete space Markov chain behaves like an SDE perturbed by small noise. In each case we can define a quasipotential by minimizing the work along paths between nodes, and in each case nodes on level sets of the quasipotential have probabilities that converge to zero at asymptotically equal rates [23]. Both methods can be applied to polystable systems to find the quasipotential associated with different basins of attraction. A consistent method for stitching together the quasipotentials to find the asymptotic behavior of the global steady state is left to future work.

7.4 Summary

In this chapter we analyzed the dynamics of nonequilibrium Markov processes using the HHD. We showed that any nonequilibrium process can be transformed into a purely rotational process if scaled by the steady state of the corresponding equilibrium process. An exact solution for the nonequilibrium steady state of a process with isolated loops was derived, and an example with linked loops was considered to illustrate how linked loops makes finding the nonequilibrium steady state a difficult problem. We then focused on analyzing steady state dynamics in a sequence of limits. These limits can be broken into two categories: limits in which forward and backward rates converge to an average rate (weak forcing), and limits in which forward and backward rates diverge (strong forcing). It was shown that, in the weak forcing limit (diffusion dominated), the steady state dynamics are governed by the HHD. In the strong forcing limit (drift dominated), the steady state dynamics are governed by a different type of potential, and that in a near deterministic limit that potential is analogous to the quasipotential used to analyze SDE's in the small noise limit.

Chapter 8

The Continuum Limit and Comparison to Quasipotentials

8.1 Preface

In Chapter 6 and Chapter 7 we applied the HHD to discrete-space continuous-time Markov chains. We showed that, in that setting the HHD has a natural physical interpretation, and is closely related to dynamics in the weak rotation/weak forcing limits. In the strong forcing limit, when diffusion is dominated by drift, a different potential was required - a quasipotential. Quasipotentials are widely used in the analysis of stochastic differential equations (SDE) in the small noise limit (cf. [23]). The objective of this chapter is to develop a consistent extension of the discrete HHD developed in Chapters 6 and 7 to the continuum so that the potential associated with the HHD can be compared to the quasipotential.

In this chapter we briefly review SDEs, and the diffusion approximations commonly

used to relate discrete-space continuous-time Markov chains to SDEs (see Section 8.2). To define a decomposition we first need to define *what* should be decomposed. On networks we decomposed the edge flow associated with the log ratio of forward and backward rates. We showed that this edge flow was closely related to a thermodynamic notion of forcing, the work to traverse paths, and the probability of observing forward and reverse trajectories. In Section 8.3 we introduce a vector field which is analogous to the edge flow. We show that it has the same thermodynamic interpretation, is related to steady states and hitting times, and for an appropriate limiting sequence of discrete-space Markov chains, the edge flow defined in Chapter 6 converges to the forces.

Next we define three potentials, the Helmholtz potential which is the scalar potential associated with the Helmholtz-Hodge Decomposition of the forces, the quasipotential, and the effective potential (see Section 8.4). It is shown that each potential is the solution to a PDE whose right hand side is associated with the forces. We show that, if the sequence of networks used to approximate $\mathbb{R}^n$ are square lattices, then the discrete potentials converge to the continuum potentials. The convergence argument leverages the spectral solution to the HHD on lattices introduced in Section 3.4. The three potentials are compared and an equivalence theorem is presented in Section 8.5. Path integral interpretations of the Helmholtz potential and the quasipotential are compared. We show that the Helmholtz potential is the large noise limit of the effective potential, and the quasipotential is the small noise limit of the effective potential. We then show that the potentials are not, in general, equivalent, but that if two of the three potentials are equivalent then all three are equivalent. We conclude by showing that for an Ornstein-Uhlenbeck process the potentials can always be chosen so that they are equivalent (see Section 8.5.2).

8.2 Stochastic Differential Equations: A Review

8.2.1 Stochastic Differential Equations

Let $X(t)$ be a stochastic process that takes values in $\mathbb{R}^n$. Then $X(t)$ is governed by a stochastic differential equation if it is a realization of the integral equation:

$$X(t) - X(0) = \int_0^t \mu(X(s))ds + B(X(s))dW(s) \tag{8.1}$$

where $\mu(x)$ is a vector field in $\mathbb{R}^n$ and defines the corresponding deterministic process: $x(t)-x(0) = \int_0^t \mu(x(s))ds$, $B(x)$ is a real, matrix valued function with n rows, and $W(t)$ is a vector of m independent Wiener processes. A Wiener process is a real-valued continuous time stochastic process with $W(0) = 0$, independent increments[1], the increments are Gaussian distributed with mean zero and variance equal to the length of the increment[2], and has continuous trajectories. Then $dW(t)$ is white noise and $W(t)$ is the accumulation of white noise over the interval $[0, t]$. A Weiner process is an example of Brownian motion. The white noise term $dW(s)$ is responsible for introducing noise to the SDE Equation (8.1). The matrix $B(x)$ is responsible for mapping from the white noise term $dW(s)$ to a multivariate random variable with covariance $B(x)B(x)^\intercal$. The matrix $D(x) = \frac{1}{2}B(x)B(x)^\intercal$ in $\mathbb{R}^{n\times n}$ is the diffusion tensor, and governs the instantaneous variance introduced to $X(t)$ by the noise term.

If the noise term is dropped then $x(t)$ obeys the ODE:

$$\frac{d}{dt}x(t) = \mu(x(t)). \tag{8.2}$$

[1] $W(s_1) - W(t_1)$ is independent of $W(s_2) - W(t_2)$ if $t_1 \leq s_1 < t_2 \leq s_2$
[2] $W(t + s) - W(t) \sim \mathcal{N}(0, s)$

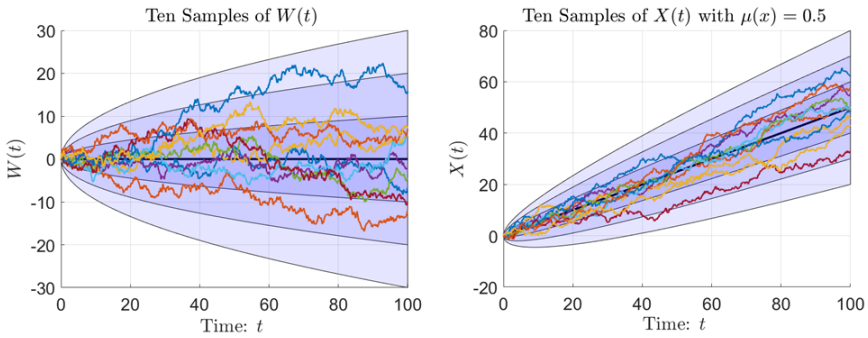

Figure 8.1: Sample trajectories $W(t)$ and $X(t)$ where $W(t)$ is a Wiener process (Brownian motion), and $X(t)$ is governed by an SDE with constant drift. The blue shaded regions represent equal one standard deviation intervals.

Thus $\mu(x)$ governs the deterministic evolution of the SDE in the limit when the noise vanishes, so is sometimes called the deterministic skeleton [23], or drift term. When we refer to drift we mean the advection of probability associated with $\mu(x)$, and when we refer to diffusion we mean the spreading of probability associated with $D(x)$ and dW.

We will usually write the SDE defined by the integral equation Equation (8.1) with the shorthand:

$$dX(t) = \mu(X(t))dt + B(X(t))dW(t). \tag{8.3}$$

Note that this integral equation is only defined if we pick a convention for performing the stochastic integral. Throughout this dissertation we will use the Itô convention, not the Stratonovich convention which is widely used in physics, since the Itô convention is non-anticipating. For a discussion of the virtues of the Itô and Stratonovich conventions see [269] and [270].

Throughout this chapter we will assume that $\mu(x)$ and $D(x)$ are Lipschitz continuous, $D(x)$ is differentiable with Lipschtiz continuous partial derivatives, and $D(x)$ is full rank for all x on the interior of the region of $x \in \mathbb{R}^n$ that can be reached by $X(t)$ with nonzero probability from any initial condition. Let Ω denote this domain.

Trajectories of an SDE can be approximated using either the Euler-Maruyama method, or the Milstein method. The Euler-Maruyama method is introduced here to clarify Equation (8.3). Let Δt be a finite time interval. Then the Euler-Maruyama approximation to $X(t)$ is defined by the recursion:

$$Y(t+\Delta t) = Y(t) + \mu(Y(t))\Delta t + B(Y(t))\Delta W, \qquad \Delta W \sim \mathcal{N}(0, \Delta t I)$$

where $Y(0) = X(0)$, ΔW is a mean zero multivariate Gaussian random variable with covariance $\Delta t\, I$, and where I is $n \times n$ the identity matrix. Note that the first two terms on the right hand side implement the standard forward Euler step for Equation (8.2). The last term adds a Gaussian distributed multivariate random variable to the Euler update. This added noise term makes $Y(t)$ a stochastic process. Thus the Euler-Maruyama update can be interpreted as a forward Euler step plus a noise term. The Euler-Maruyama method converges with strong order $1/2$ in Δt. That is, in one-dimension the expected error $\mathbb{E}\left[|Y(T) - X(T)|^{1/2} \leq C(\Delta t)^{1/2}\right]$ for $\Delta t < 1$, any time T, and some constant C [271]. [3]

Let $p(x,t)$ be the probability density that $X(t) = x$. Then if $X(t)$ obeys the SDE defined by Equation (8.3), the density $p(x,t)$ satisfies the Fokker-Planck equation [240, 24]:

$$\partial_t p(x,t) = -\sum_{i=1}^{n} \partial_{x_i}[\mu_i(x)p(x,t)] + \sum_{i,j=1}^{n} \partial^2_{x_i x_j}[D_{ij}(x)p(x,t)] \tag{8.4}$$

where $D(x) = \frac{1}{2}B(x)B(x)^\intercal$ is the diffusion tensor. The Fokker-Planck equation is an

[3] The Milstein method modifies the Euler-Maruyama method, and converges faster in Δt (first order convergence instead of convergence to order $1/2$) [271]. In one dimension the Milstein update is:

$$Y(t+\Delta t) = Y(t) + \mu(Y(t))\Delta t + B(Y(t))\Delta W + \frac{1}{2}B(Y(t))\frac{d}{dx}B(Y(t))\left((\Delta W(t))^2 - \Delta t\right).$$

In higher dimensions the Milstein update can be defined similary, where the last term is replaced with a tensor product between $B(x)$, its first order partials, and the tensor with entries defined by the stochastic integrals $\int_t^{t+\Delta t}(W_k(t+s) - W_k(t))dW_l(s)$ [271].

advection-diffusion equation. The differential operator acting on $p(x,t)$ on the right hand side of the Fokker-Planck equation is the forward Kolmogorov operator. The forward Kolmogorov operator is responsible for governing the flow of probability forward in time, so is analogous to the Laplacian L that governs the flow of probability for the discrete space processes considered in Chapter 6 and Chapter 7. The Fokker-Planck equation can also be written as the divergence of a flux by writing:

$$
\begin{aligned}
&-\sum_{i=1}^{n} \partial_{x_i}[\mu_i(x)p(x,t)] = -\nabla \cdot [\mu(x)p(x,t)] \\
&\sum_{i,j=1}^{n} \partial^2_{x_i x_j}[D_{ij}(x)p(x,t)] = \sum_{i=1}^{n} \partial_{x_i} \left(\left(\sum_{j=1}^{n} \partial_{x_j} D_{ij}(x) \right) p(x,t) + \sum_{j=1}^{n} D_{ij}(x) \partial_{x_j} p(x,t) \right)
\end{aligned}
$$

and expanding the diffusive term:

$$
\sum_{i,j=1}^{n} \partial^2_{x_i x_j}[D_{ij}(x)p(x,t)] = \nabla \cdot \left(\sum_{j=1}^{n} (\partial_{x_j} D_j(x)) p(x,t) + D(x) \nabla p(x,t) \right)
$$

where $D_j(x)$ denotes the j^{th} column of the diffusion tensor. Note that the tensor is symmetric so $D_j(x)$ is also the j^{th} row.

Then, grouping terms:

$$
\begin{aligned}
\partial_t p(x,t) &= -\nabla \cdot J(x,t) \\
J(x,t|p) &= [\mu(x) - \nabla \cdot D(x)]\, p(x,t) - D(x) \nabla p(x,t)
\end{aligned}
\tag{8.5}
$$

where $J(x,t|p)$ is the probability flux, and $\nabla \cdot D(x)$ is interpreted as the divergence of each row.

The probability flux is a combination of the two terms. The first (bracketed) term is advective, and is responsible for the drift of probability. Note that advection is influenced

both by the deterministic term $\mu(x)$, and by the divergence of each row of the diffusion tensor. If a row of the diffusion tensor is diverging at some x, then the noise has a larger effect at that x than at neighboring states, so probability distribution tends to move away from x. The advective term governs the expected motion of $X(t)$. The second term is diffusive, and governs the spread of probability. If $X(0) = x_0$ is fixed then $p(x, 0) = \delta(x - x_0)$ and the instantaneous rate of change in the variance is governed by the second term.

The Fokker-Planck equation can be used to find the time evolution of any observable of the process. Here we will use the Fokker-Planck equation to derive the backward equation that governs the time evolution of the first two moments. This derivation is meant to provide a clearer appreciation for the roles of μ and D, and to link the drift and diffusion terms to the instantaneous time evolution of the moments. Similar evolution equations can be derived in discrete space given the transition rates. Then, by choosing the drift and diffusion terms appropriately it is possible to construct an SDE with the same time evolution equations for the first and second moments as a discrete space process. This link will be used to motivate our definition of forces in continuous space by relating them to the edge flow in discrete space.

First, let $g(x)$ be a continuously differentiable scalar-valued function of x, and let Ω be a finite domain. We require that $g(x)$ is defined for all $x \in \Omega$. Then:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[g(X(t))] &= \frac{d}{dt}\int_\Omega g(x)p(x,t)d\Omega \\ &= \int_\Omega g(x)\left[\partial_t p(x,t)\right]d\Omega = -\int_\Omega g(x)\left[\nabla \cdot J(x,t|p)\right]d\Omega\end{aligned}$$

Then, using Green's first identity (integration by parts):

$$\frac{d}{dt}\mathbb{E}[g(X(t))] = \int_\Omega [\nabla g(x)] \cdot J(x,t|p) d\Omega - \int_\Gamma g(x) \left[J(x,t|p) \cdot \hat{n}(x)\right] d\Gamma$$

where $\Gamma = \partial\Omega$ is the boundary of the domain. By assumption the process $X(t)$ cannot leave the domain, so the probability flux through the boundary, $[J(x,t|p) \cdot \hat{n}(x)]$, must always be zero. Therefore, the time evolution of any observable is given by an integral over the domain of the gradient of the function that defines the observable dotted into the probability flux:

$$\frac{d}{dt}\mathbb{E}[g(X(t))] = \int_\Omega [\nabla g(x)] \cdot J(x,t|p) d\Omega. \tag{8.6}$$

Equation (8.6) can be read as the change in the observable due to the accumulated change in the observable associated with the flux of probability at each point in the domain. The same equation is valid for a discrete-space continuous-time Markov chain. For a discrete-space Markov chain:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[g(X(t))] &= \sum_{x\in\Omega} g(x)\frac{d}{dt}p(x,t) = -\sum_{x\in\Omega} g(x)[-G^\intercal J(x,t|p)] \\ &= g^\intercal G^\intercal J(p) = [Gg]^\intercal J(p).\end{aligned}$$

where Gg is the gradient of the vector g with entries $g(x)$.

Substitute Equation (8.5) into Equation (8.6). Then:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[g(X(t))] &= \int_\Omega [\nabla g(x)] \cdot J(x,t|p) d\Omega \\ &= \int_\Omega [\nabla g(x)] \cdot \left([\mu(x) - \nabla \cdot D(x)]\, p(x,t) - D(x)\nabla p(x,t)\right) d\Omega \\ &= \mathbb{E}\left[\nabla g(X) \cdot (\mu(X) - \nabla \cdot D(X))\right] - \int_\Omega [\nabla g(x)] \cdot \left(D(x)\nabla p(x,t)\right) d\Omega.\end{aligned}$$

To simplify note that $[\nabla g(x)] \cdot (D(x)\nabla p(x,t)) = [\nabla g(x)]^{\intercal} D(x) (\nabla p(x,t))$ so equals $[D(x)\nabla g(x)] \cdot (\nabla p(x,t))$. Then, applying Green's identity:

$$\int_{\Omega} [\nabla g(x)] \cdot (D(x)\nabla p(x,t))\, d\Omega = -\int_{\Omega} [\nabla \cdot D(x)\nabla g(x)]\, p(x,t) d\Omega + \int_{\Gamma} (\nabla p(x,t))\, [D(x)\nabla g(x)] \cdot \hat{n}(x) d\Gamma$$

but $D(x)\nabla g(x)$ must be in the range of the diffusion tensor, which must be orthogonal to the boundary of the domain, otherwise it would be possible for $X(t)$ to leave the domain. Therefore the boundary term vanishes leaving:

$$\frac{d}{dt}\mathbb{E}[g(X(t))] = \mathbb{E}\left[\nabla g(X) \cdot (\mu(x) - \nabla \cdot D(x)) + \nabla \cdot D(X)\nabla g(X)\right]$$

To complete the calculation apply the product rule to the last term in the expectation:

$$\begin{aligned}
\nabla \cdot D(x)\nabla g(x) &= \sum_{i=1}^{n} \partial_{x_i} \left[\sum_{j=1}^{n} d_{ij}(x)\partial_{x_j} g(x)\right] \\
&= \sum_{i,j=1}^{n} (\partial_{x_i} d_{ij}(x))(\partial_{x_j} g(x)) + \sum_{i,j=1}^{n} d_{ij}(x)\partial^2_{x_i x_j} g(x) \\
&= (\nabla \cdot D(x)) \cdot \nabla g(x) + \langle D(x), H(x)\rangle
\end{aligned}$$

where $\langle A, B\rangle = \sum_{i,j} A_{ij}B_{ij}$ is the matrix inner product, and $H(x)$ is the Hessian ($h_{ij}(x) = \partial^2_{x_i x_j} g(x)$. Then, the inner product between the divergence of $D(x)$ and the gradient of $g(x)$ cancels with the matching term in the time derivative of the observable. Therefore, the time

derivative of the observable is the expectation:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[g(X(t))] &= \mathbb{E}\left[\mu(X(t))\cdot\nabla g(X(t)) + \langle D(X(t)), H(X(t))\rangle\right] \\ &= \int_\Omega \left[\sum_{i=1}^n \mu_i(x)\partial_{x_i} g(x) + \sum_{i,j=1}^n d_{ij}(x)\partial^2_{x_i x_j} g(x)\right] p(x,t)d\Omega.\end{aligned} \tag{8.7}$$

The differential operator acting on $g(x)$ inside the expectation is the backward Kolmogorov operator. The backward Kolmogorov operator is the adjoint to the Kolmogorov operator, and is analogous to the transpose of the Laplacian which appears in the master equation. It governs the time evolution of the probability that $X(t) = x$ given that $X(t + s) = y$ for some fixed final state y. In other words, it governs the backwards time evolution of probability. The forward Kolmogorov operator answers the question, "where will I be?" The backward Kolmogorov operator answers the question, "where was I?"

It follows immediately from Equation (8.7) that if $X(0) = x_0$ then $p(x,t) = \delta(x - x_0)$ so the instantaneous time evolution of any observable given a fixed initial condition is:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[g(X(t))|X(0) = x]|_{t=0} &= \mu(x)\cdot\nabla g(x) + \langle D(x), H(x)\rangle \\ &= \sum_{i=1}^n \mu_i(x)\partial_{x_i} g(x) + \sum_{i,j=1}^n d_{ij}(x)\partial^2_{x_i x_j} g(x)\end{aligned} \tag{8.8}$$

so the backward Kolmogorov equation also governs the instantaneous rate of change of any observable given a fixed initial condition.

Equation (8.7) can be used to compute the equations governing the time evolution of the expected state and covariance. For the expected state set $g(x) = x_i$ for each i. For the covariance set $g(x) = x_i x_j$ for each pair i, j, then subtract off the time evolution of the expected state squared. Let $\bar{g}(t) = \mathbb{E}[g(X(t))]$ denote the expected value of a function

$g(X)$ at time t. Then, substituting into Equation (8.7) yields the evolution equations:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= \mathbb{E}[\mu(X(t))] \\ \frac{d}{dt}\mathbb{V}(t) &= \mathbb{E}[(X(t)-\bar{x}(t))\mu(X(t))^\intercal + \mu(X(t))(X(t)-\bar{x}(t))^\intercal] + \mathbb{E}[D(X(t))].\end{aligned} \tag{8.9}$$

where $\mathbb{V} = \mathbb{E}[(X(t)-\bar{x}(t))(X(t)-\bar{x}(t))^\intercal]$ is the covariance in $X(t)$.

Therefore the rate of change in the expected state is the expected value of the drift term, and the expected rate of change in the covariance is a combination of two terms associated with the covariance in the state and drift term, and the diffusion tensor. The expectation $\mathbb{E}[(X(t)-\bar{x}(t))\mu(X(t))^\intercal] = \mathbb{E}[(X(t)-\bar{x}(t))(\mu(X(t)))-\bar{\mu}(t))^\intercal]$ is the covariance between $X(t)$ and the drift term $\mu(X(t))$ that drives the deterministic version of the SDE.

Since $p(x,0) = \delta(x-x_0)$ the instantaneous rate of change in the expected state and covariance is:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)]|_{t=0} &= \mu(x_0) \\ \frac{d}{dt}\mathbb{E}[(X(t)-\bar{x}(t))(X(t)-\bar{x}(t))^\intercal]|_{t=0} &= 2D(x_0).\end{aligned} \tag{8.10}$$

Therefore $\mu(x)$ can be interpreted as the instantaneous change in the expected state given an initial condition, and $D(x)$ can be interpreted as the instantaneous covariance produced given an initial state.

8.2.2 Diffusion Approximations and the Continuum Limit

Stochastic differential equations are widely used to approximate discrete-space continuous-time Markov processes. Here we review the standard techniques for constructing a diffusion approximation to a discrete-space continuous-time Markov process. We focus on reaction networks since there is a natural limiting procedure for reaction networks that converges to an SDE. This limit will establish a link between SDEs and discrete-space

continuous-time Markov chains that will be used to relate our analysis of SDEs to our analysis of discrete-space Markov chains.

Suppose that $Y(t)$ is a discrete-space continuous-time Markov process defined by a reaction network with reactions $\mathcal{R}$, propensity functions $\lambda_k(x)$, and stoichiometry vectors s_k. Then the probability that $Y(t) = y$, written $p(y,t)$, is governed by the master equation:

$$\frac{d}{dt}p(y,t) = \sum_{r_k \in \mathcal{R}} (\lambda_k(y - s_k)p(y - s_k, t) - \lambda_k(y)p(y,t)).$$

Assume that the reactions are grouped so that each reaction has a unique stoichiometry vector. That is, all reactions with the same stoichiometry vector are lumped into a single reaction with propensity equal to the sum of the propensities. Then the state space can be represented as a graph with edges associated with each reaction.

The time evolution of any moment can be expressed by the generic equation:

$$\frac{d}{dt}\mathbb{E}[g(Y(t))] = \mathbb{E}\left[\sum_{r_k \in \mathcal{R}} \lambda_k(Y)\left(g(Y + s_k) - g(Y)\right)\right].$$

Then the expected state and covariance are governed by the evolution equations:

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}[Y(t)] &= \mathbb{E}\left[\sum_{r_k \in \mathcal{R}} \lambda_k(X(t))s_k\right] = \mathbb{E}[S\lambda(Y(t))] \\
\frac{d}{dt}\mathbb{V}(t) &= \mathbb{E}\left[\sum_{r_k \in \mathcal{R}} s_k\lambda_k(Y(t))(Y(t) - \bar{y}(t)^\intercal + (Y(t) - \bar{y}(t))(s_k\lambda_k(Y(t)))^\intercal + s_k^\intercal s_k \bar{\lambda}_k(t)\right]
\end{aligned} \tag{8.11}$$

where $S \in \mathbb{R}^{n \times |\mathcal{R}|}$ is the stoichiometry matrix with columns equal to the stoichiometry vectors, and $\lambda(x) \in \mathbb{R}^{|\mathcal{R}|}$ is the vector with entries equal to the propensities of each reaction, and $\bar{\lambda}_k(t) = \mathbb{E}[\lambda_k(Y(t))]$ is the expected propensity of reaction r_k.

Let:

$$\begin{aligned}\mu(y) &= S\lambda(y) = \sum_{r_k \in \mathcal{R}} s_k \lambda_k(y) \\ B(y) &= S\,\mathrm{diag}\left(\sqrt{2\lambda(y)}\right), \quad D(x) = \frac{1}{2}B(x)B(x)^\intercal = \sum_{r_k \in \mathcal{R}} s_k s_k^\intercal \lambda_k(y).\end{aligned} \tag{8.12}$$

If $X(t)$ is an SDE governed by $dX = \mu(X)dt + B(X)dW$, then the instantaneous change to the expected state and covariance given any initial condition is the same as the instantaneous change to the expected state and covariance in $Y(t)$ given any initial condition. Matching the instantaneous change to the expected state and covariance produces the Langevin approximation $X(t)$ to the discrete-space continuous-time process $Y(t)$ [242]:

$$\begin{aligned}\mu(y) &= S\lambda(y) = \sum_{r_k \in \mathcal{R}} s_k \lambda_k(y) \\ B(y) &= S\mathrm{diag}\left(\sqrt{2\lambda(y)}\right), \quad D(x) = \frac{1}{2}B(x)B(x)^\intercal = \sum_{r_k \in \mathcal{R}} s_k s_k^\intercal \lambda_k(y) \\ X(0) &= Y(0), \quad dX(t) = \mu(X(t))dt + B(X(t))dW(t)\end{aligned} \tag{8.13}$$

where $W(t)$ is a Wiener process with $|\mathcal{R}|$ independent increments, one for each reaction.

When $Y(t)$ represents a reaction system it is often possible to introduce a system size parameter. For example, if $Y(t)$ is the number of chemical species of certain types in solution, then the volume of the solution is a system size parameter, and $Y(t)$ can be replaced with a vector $Z(t)$ representing concentration. Then, in the large system size limit $Z(t)$ converges to the Langevin approximation to $Y(t)$ rescaled by the system size parameter [240]. Typically in the large system size limit the diffusion tensor vanishes proportional to one over the system size, so in the large system size limit the process becomes closer to deterministic and the noise becomes increasingly small [24]. The Langevin approximation

can also be derived by using a tau-leaping approximation, in which it is assumed that many reactions occur per time step tau, but that the reactions have a small impact on the state, so the propensities do not change significantly over the course of the time interval. Then the number of reactions of each type is, approximately, a Poisson random variable, which if approximated with a Gaussian with the same mean and variance, recovers the Langevin approximation as τ goes to zero [242].

Equation (8.13) establishes a link between the transition rates of a discrete-space continuous-time Markov chain and the drift and diffusion terms in an SDE. In this context the SDE is an approximation to the discrete-space process. Diffusion approximations are widely used, although in some cases there are nontrivial convergence issues (cf. [272, 273]).

The key first step in our analysis of discrete-space continuous-time Markov chains was to consider the geometric difference and geometric average of each pair of forward and backward transition rates. The principle of microscopic reversibility guarantees that for any forward reaction there is a reverse reaction. The reverse reaction may be highly unlikely, so may be ignored in some models. Suppose that the reverse reactions are not ignored. Then the reactions can be grouped in pairs. All reactions that have stoichiometry s_k are grouped into a single forward reaction, and all reactions that have stoichiometry $-s_k$ are grouped into the reverse reaction. Let $\lambda_k^+(y)$ denote the rate of the forward reactions and $\lambda_k^-(y)$ denote the reverse rate. Then each pair of reactions corresponds to a class of edges in the network. Let $\mathcal{E}$ be the set of undirected edges connected to each state. Then:

$$\begin{aligned} \mu(y) &= \sum_{k\in\mathcal{E}} s_k(\lambda_k^+(y) - \lambda_k^-(y)) \\ D(y) &= \sum_{k\in\mathcal{E}} s_k s_k^T(\lambda_k^+(y) + \lambda_k^-(y)). \end{aligned} \tag{8.14}$$

Therefore the drift term is determined by the arithmetic difference in the forward and

backward rate, and the diffusion term is associated with the arithmetic average of the forward and backward rate. This observation is familiar in the study of birth-death processes, where the difference in birth and death rates gives the growth rate which determines the deterministic growth of a population, and the sum of birth and death rates controls the intensity of demographic stochasticity.

Suppose that the reaction network is modeling a birth-death process involving n different populations. Suppose that the only events which change each population are the birth of a new member, or the death of an existing member. Then each transition changes only one species, so the stoichiometry vectors are all canonical basis vectors. Let $\lambda_k^+(y)$ be the birth rate of species k given populations y, and $\lambda_k^-(y)$ be the death rate. Then $s_k = e_k$ and:

$$\begin{aligned} \mu_k(y) &= \lambda_k^+(y) - \lambda_k^-(y) \\ D(y) &= \text{diag}(\lambda_k^+(y) + \lambda_k^-(y)). \end{aligned} \tag{8.15}$$

The corresponding network is the square lattice $\mathbb{Z}^n$. Therefore, for a birth-death process the discrete space model is a random walk on a square lattice, and the difference in birth and death rates in species k determines the k^{th} entry of the drift term μ, and the sum of the birth and death rates in species k determines the k^{th} diagonal entry of the diffusion term D. Thus, for a birth-death process, there is simple mapping between the transition rates of the discrete-space process and the drift and diffusion term.

The fact that μ is usually related to a difference in forward and backward rates, and D is related to their sum is the key observation that will help us link our analysis of discrete-space processes to SDEs. We will pay special attention to square lattices since they offer the easiest link between discrete-space processes and SDEs.

8.3 Forces

In order to define a decomposition we first have to choose what to decompose. For discrete-space continuous-time Markov chains, we decomposed the edge flow f where f_{ij} was the log of the ratio of the forward and backward transition rates between states i and j. We showed that this edge flow was analogous to work, or, if scaled by the edge length, the average force acting on the system as it crossed the edge. Then the mean rate at which the system dissipated heat into the environment was controlled by the inner product of the probability flux with the edge flow (see Section 6.5).

An edge flow on a network is analogous to a vector field in the continuum, so it is reasonable to start by looking for a vector field that has the same thermodynamic interpretation as the edge flow. Qian et al. propose the vector field $D^{-1}(x)\mu(x)$ [20]. This vector field is analogous to force in a thermodynamic system, and its inner product with the probability fluxes controls the rate at which the system dissipates heat into the environment [20]. Qian proposes applying a Helmholtz-Hodge decomposition to separate the forces into a conservative and circulating component [20, 239, 274]. If this field is conservative then the corresponding SDE obeys detailed balance, is time-reversible, and does not produce entropy at steady state [20]. Thus the field $D^{-1}(x)\mu(x)$ is a natural candidate to decompose.

Before advancing further it is worth asking, why do the forces take this form? The forces are large at x if the drift term is larger than the diffusion term at x. In particular, $f(x)$ is large if the probability of sampling a noise vector dW such that BdW matches the drift term is small. Therefore, the forces are large where the drift term dominates diffusion. In that case small transitions in the direction of $f(x)$ are close to irreversible. In contrast, if diffusion dominates drift then small transitions are easily reversed. As shown before, the probability of forward transitions relative to backward transitions is associated with the

work needed to make the associated transition. When a particular transition exchanges a large amount of heat with the environment (requires a large amount of work) it will be close to irreversible, so the drift in the direction of the transition should be larger than the diffusive term. In this context the assumption that $D(x)$ is invertible is analogous to the reversibility assumption used throughout Chapter 6 and Chapter 7.

We will make one important modification to this field. Instead of working with $D^{-1}\mu(x)$ we will work with the vector field:

$$f(x) = D^{-1}(x)\left(\mu(x) - \nabla \cdot D(x)\right). \tag{8.16}$$

This difference in definition arises from our choice to use the Itô convention. Qian uses the Stratonovich convention. Note that subtracting $\nabla \cdot D(x)$ from $\mu(x)$ is the standard transformation needed to move between the Itô form of the Fokker-Planck equation and the Stratonovich form [269]. Therefore these two definitions of the force are consistent up to the choice of convention for stochastic integration.

The choice to decompose the field $f(x)$ defined by Equation (8.16) can also be motivated without a thermodynamic interpretation. Consider a discrete-space process again. If the process obeyed detailed balance then, for the right choice of edge flow, the edge flow would be conservative, and the corresponding potential would uniquely determine the steady state distribution via the Boltzmann equation. Conversely, we should choose the forces so that, if the forces are not conservative, then the system should not obey detailed balance. Therefore, if we wish to find a vector field which is the natural analog to the edge flow we studied in Chapter 6 and Chapter 7, then we should seek a vector field that is conservative if and only if the SDE obeys detailed balance, and whose potential equals the log of the steady state when in detailed balance.

To start, consider the one-dimensional case. Then the Fokker-Planck equation (Equation (8.4)) reads:

$$\partial_t p(x,t) = -\partial_x[\mu(x)p(x,t)] + \partial_x^2[d(x)p(x,t)] = -\partial_x\left(\mu(x)p(x,t) - \partial_x d(x)p(x)\right).$$

At steady state the distribution $p(x,t)$ must stop changing. Therefore, if $q(x)$ denotes the steady state it must satisfy:

$$-\partial_x\left(\mu(x)q(x) - \partial_x d(x)q(x)\right) = 0.$$

Therefore there must be some constant C such that $\mu(x)q(x) - \partial_x d(x)q(x) = C$. Use the product rule to move the diffusion term, $d(x)$, outside of the derivative. Then:

$$(\mu(x) - \partial_x d(x))q(x) - d(x)\partial_x q(x) = C.$$

To find the constant note that the term, $(\mu(x) - \partial_x d(x))q(x) - d(x)\partial_x q(x)$, is the probability flux. The only way for the probability flux to be nonzero, but not changing, is if there is a constant probability flux for all x. This is impossible if the domain is bounded, since probability cannot flow through the boundary. If the domain is unbounded, or if the domain has periodic boundaries then the flux could equal a constant everywhere. We are looking for a solution in detailed balance, so the flux must be zero at steady state, so $C = 0$. Then the steady state equation is the separable differential equation:

$$d(x)\partial_x q(x) = (\mu(x) - \partial_x d(x))q(x).$$

We assumed that the diffusion tensor is always invertible so $d(x) > 0$ for all x in the

domain. Divide across by $d(x)$. Then:

$$\partial_x q(x) = \frac{\mu(x) - \partial_x d(x)}{d(x)} q(x) = f(x) q(x)$$

where $f(x)$ are the forces defined by Equation (8.16).

Let:

$$S(x) = -\int_{x_0}^{x} f(y) dy \tag{8.17}$$

for some initial x_0. Then $-\partial_x S(x) = f(x)$ so $S(x)$ plays the role of a potential.[4] Then, the steady state is given by:

$$q(x) = \frac{1}{Z} \exp\left(\int_{x_0}^{x} f(y) dy\right) = \frac{1}{Z} \exp(-S(x)) \tag{8.18}$$

where Z is the necessary normalization.

Therefore, under the requirement that the steady state flux is zero (detailed balance), then, in one-dimension, the potential function $S(x)$ defined so that $-\nabla S(x) = f(x)$ where $f(x)$ are the forces given by Equation (8.16) satisfies a Boltzmann type relation with the steady state. Thus, in one-dimension and in detailed balance, $f(x)$ plays the same role as the edge flow used for discrete-space processes, and $S(x)$ plays the same role as the scalar potential ϕ.

The same observation about $f(x)$ and the steady state in detailed balance extends to n dimensional processes. In detailed balance the steady state flux must be zero everywhere, which requires:

$$J(x|q) = (\mu(x) - \nabla \cdot D(x))\, q(x) - D(x) \nabla q(x) = 0.$$

[4] The potential can be simplified by noting that $\int_{x_0}^{x} \frac{d'(y)}{d(y)} dy = \log(d(x)/d(x_0))$. Therefore $S(x) = -\int_{x_0}^{x} \mu(x)/d(x) + \log(d(x)/d(x_0))$ and $\exp(-S(x)) = d(x_0)/d(x) \exp(-\int_{x_0}^{x} \mu(x)/d(x))$.

Rearranging, and multiplying on both sides by $D^{-1}(x)$ yields the steady state equation:

$$\nabla q(x) = D^{-1}(x)(\mu(x) - \nabla \cdot D(x))q(x) = f(x)q(x).$$

Now let $S(x) = -\log(q(x))$. Then $q(x) = \exp(-S(x))$ so $\nabla q(x) = \nabla \exp(-S(x)) = [-\nabla S(x)] \exp(-S(x))$ or, more succinctly, $\nabla q(x) = [-\nabla S(x)]q(x)$. Then the steady state equation reads:

$$[-\nabla S(x)]q(x) = f(x)q(x) \tag{8.19}$$

so, for $x \in \text{supp}(q(x))$, the forces $f(x)$, and steady state $q(x)$, are related by the potential function $S(x)$ such that:

$$-\nabla S(x) = f(x), \quad q(x) \propto \exp(-S(x)). \tag{8.20}$$

Equation (8.20) shows that, if the process obeys detailed balance (no flux at steady state), then the steady state distribution obeys a Boltzmann type relationship with the potential $S(x)$ such that $-\nabla S(x) = f(x)$. Note that this equation only makes sense if $f(x)$ is conservative, otherwise there is no $S(x)$ such that $-\nabla S(x) = f(x)$. As an immediate consequence, if $f(x)$ is not conservative on the support of the steady state ($f(x)$ could be rotational off the support if the support is absorbing) then there is no steady state distribution such that the steady state fluxes are all zero since there is no solution to Equation (8.19). Thus $f(x)$ plays the same role as the edge flow in our original theory. If $f(x)$ is conservative, then it is possible to solve Equation (8.19), so there exists a steady state of the form Equation (8.20) for which there is no steady state flux, thus the process has a true equilibrium and obeys detailed balance. If $f(x)$ is not conservative, then, provided there is not an absorbing set where $f(x)$ is conservative, there is no steady state such that

all of the steady state fluxes vanish, thus the process does not obey detailed balance. For a more general discussion of the relationship between $f(x)$ and detailed balance see [20].

These conclusions strongly motivate a potential decomposition of $f(x)$. If it is possible to find a potential such that $f(x) = -\nabla S(x)$, then the process obeys detailed balance and the steady state is determined by $S(x)$, and if it is not possible to find such a potential then the process is a nonequilibrium process. More precisely (see Theorem 1 in [20]):

Lemma 37 (Detailed Balance for SDEs)**.** *If $X(t)$ is a stationary stochastic process governed by the SDE $dX = \mu(X)dt + B(X)dW$, where $\mu(x)$ is differentiable, $B(x)$ is twice differentiable, and $D(x) = \frac{1}{2}B(X)B(X)^T$ is invertible for all x that could possibly be reached by $X(t)$, then $X(t)$ obeys detailed balance if and only if the forces $f(x) = D^{-1}(x)(\mu(x) - \nabla \cdot D(x))$ are conservative. If the forces are conservative then $X(t)$ is time-reversible [20].*

The forces $f(x)$ and associated potential (if it exists) are also closely related to first passage time problems [23, 275, 276], and problems that involve the backward operator [24]. We used this fact in [276] to approximate the asymptotic behavior of mean first passage times to extinctions analytically.

8.3.1 Forces and Edge Flows: The Continuum Limit

It remains to show that the forces defined by Equation (8.16) are consistent with the edge flow defined by Equation (6.3) that were used throughout Chapter 6 and Chapter 7. We have shown that the forces have the same thermodynamic interpretation, and, like the edge flow, control whether or not $X(t)$ obeys detailed balance. We have not shown that the edge flow converges to the forces in an appropriate continuum limit.

The section proceeds as follows. First, the generic relationship between the forces and the edge flow is identified using the observations highlighted at the end of the section on the Langevin approximation. Then two specific examples are provided where the edge flow converges to the forces. First, a simple family of biased random walks are introduced, and it is shown that, as the jump size is shrunk to zero, the edge flow converges to the forces. Then, it is shown that if a family of discrete processes are chosen to discretize an SDE, then the edge flow of the discrete approximations converge to the forces of the original SDE.

The key link between the continuum and discrete-space continuous-time processes was outlined in Section 8.2.2 where we showed that the difference in forward and backward rates determines the drift term $\mu(x)$, and the sum determines the diffusion term $D(x)$. The geometric difference in forward and backward rates determined the edge flow f, while the geometric average of the forward and backward rates determined the conductances ρ. So, speaking roughly, the flow is related to the drift term, and the conductances are related to the diffusion term. More precisely, suppose we have a pair of rates l_+ and l_-. Then $l_+ = \rho \exp(f)$ and $l_- = \rho \exp(-f)$, so $l_+ - l_- = 2\rho \sinh(f)$ while $l_+ + l_- = 2\rho \cosh(f)$. Therefore:

$$\frac{l_+ - l_-}{l_+ + l_-} = \tanh(f).$$

The numerator is associated with μ and the denominator is associated with D. Thus the hyperbolic tangent of the edge flow is related to the ratio of drift over diffusion. The forces were $D^{-1}(\mu - \nabla \cdot D)$ so are also a ratio of drift to diffusion. If the edge flow is small then $\tanh(f) = f + \mathcal{O}(f^3)$ so the edge flow is approximately equal to the ratio of drift to diffusion, and thus to the forces. In order for a discrete space process to converge to a continuum process the edge flow typically needs to converge to zero, as in a hydrodynamic limit of a random walk with parabolic scaling where the forward and backward rates

converge to each other as the discretization is refined.

For example, consider a discrete time random walk on a sequence of nodes that discretize $\mathbb{R}$. Let x_i denote the position of the i^{th} node, with $x_i = i\Delta x$ and where $i \in -\mathbb{Z} \cup \mathbb{Z}$. Let $t_j = j\Delta t$ where $j \in \mathbb{Z}$. Then let the time interval and space interval satisfy a parabolic scaling of the form:

$$\Delta t = \frac{\Delta x^2}{2d}.$$

for some positive number d.

Assume that at any time t_j the probability that a walker at x_i moves to the right is given by $l_+ = \frac{1}{2}(1+\epsilon)$ and the probability that the walker moves left is $l_- = 1 - l_- = \frac{1}{2}(1-\epsilon)$. In addition assume that ϵ scales with Δt so that:

$$\epsilon = \frac{\mu \Delta x}{2d}.$$

for some $\mu \in \mathbb{R}$.

Let $p_{i,j} = p(x_i, t_j)$ be the probability that a walker occupies the i^{th} state at time t_j. Then $p_{i,j}$ obeys the difference equation:

$$p_{i,j} = l_+ p_{i-1,j-1} + l_- p_{i+1,j-1} = \frac{1}{2}\left[(1+\epsilon)p_{i-1,j-1} + (1-\epsilon)p_{i+1,j-1}\right]$$

Subtracting $p_{i,j-1}$ from both sides and rearranging gives:

$$p_{i,j} - p_{i,j-1} = \frac{1}{2}\left[p_{i-1,j-1} - 2p_{i,j-1} + p_{i+1,j-1}\right] - \frac{\epsilon}{2}\left[p_{i+1,j-1} - p_{i-1,j-1}\right].$$

Next, divide both sides by Δt and substitute in for ϵ:

$$\frac{1}{\Delta t}\left[p_{i,j} - p_{i,j-1}\right] = \frac{d}{\Delta x^2}\left[p_{i-1,j-1} - 2p_{i,j-1} + p_{i+1,j-1}\right] - \frac{\mu}{2\Delta x}\left[p_{i+1,j-1} - p_{i-1,j-1}\right].$$

Notice that the left hand side is a difference approximation to a time derivative, the first term on the right hand side is the central difference approximation to a second order spatial derivative, and the last term is a central difference approximation to a first order spatial derivative [76]. In the limit that Δx, and as a consequence Δt, goes to zero the first term is a time derivative, the second term is a second order spatial derivative, and the last term is a first order spatial derivative. Therefore the difference equation converges to a advection-diffusion PDE:

$$\partial_t p(x,t) = -\mu \partial_x f(x,t) + d\partial_x^2 f(x,t)$$

which has the same form as the Fokker-Planck equation for the SDE $dX = \mu dt + \sqrt{2d} dW$.

Returning to the original transition probabilities, when Δt is small l_+ and l_- converge to Δt times the instantaneous transition rates between states since the continuous time process has exponentially distributed event times [244]. Thus, for small Δt, l_+/l_- converges to the ratio of forward and backward rates. Therefore, in the limit as Δt goes to zero, the edge flow is:

$$f = \frac{1}{2}\log(l_+/l_-) = \frac{1}{2}\log\left(\frac{\frac{1}{2}(1+\epsilon)}{\frac{1}{2}(1-\epsilon)}\right) = \epsilon + \mathcal{O}(\epsilon^3) = \frac{\mu}{d}\frac{\Delta x}{2} + \mathcal{O}(\Delta x^3).$$

The ratio μ/d is the force for the SDE achieved in the limit, therefore the edge flow converges to $\Delta x/2$ times the force. The factor of two, as usual, comes from our choice to work with the square root of the ratio of forward and backward rates, so has no real significance outside of convention.

Thus, for some sequences of biased random walks in one-dimension with appropriately chosen scaling, the edge flow on the network converges to the forces defined by Equation (8.16). Our goal is to show that this convergence holds for more general sequences

of networks. Here we show that, if we pick a sequence of networks with transition rates designed to ensure convergence to a particular SDE, then the edge flow converges to the forces.

Suppose that we start with an SDE, and approximate it with a discrete-space continuous-time process. Consider an SDE defined on the real line $x \in \mathbb{R}$. The corresponding Fokker-Planck equation is:

$$\frac{d}{dt}\pi(x) = -\partial_x(\mu(x)\pi(x)) + \partial_x^2(D(x)\pi(x)).$$

Here $\pi(x)$ is used instead of $p(x)$ since the Fokker-Planck equation governs the evolution of a probability density, which we will approximate with a set of probabilities p on nodes.

The Fokker-Planck equation is a one-dimensional PDE. We would like to approximate this PDE with a master equation of the form:

$$\frac{d}{dt}p = Lp.$$

This is a classic problem in numerical differential equations. There are a variety of standard methods for discretizing advection-diffusion type equations (see [76]). Here we will focus on a second order finite volume method designed to conserve probability. In addition we require that the discretized differential operator L is interpretable as a Laplacian for a continuous-time Markov chain. This requires that all the off-diagonal entries are nonnegative, and all the columns sum to zero. Finally, we would like the associated network to reflect the ordering of the real line, so require that each node is connected exclusively to its nearest neighbors. This requires that L is tridiagonal. An appropriate discretization scheme follows.

First, discretize the real line into a series of evenly spaced cells width Δx. Denote the

center of the j^{th} cell x_j and let $x_j = j\Delta x$. The boundaries between the cells are $x_j \pm \frac{1}{2}\Delta x$. As short hand these will be denoted $x_{j\pm\frac{1}{2}}$.

Let $p_j(t)$ represent the probability that the system is in the j^{th} cell at time t. Then:

$$p_j = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \pi(x)dx. \tag{8.21}$$

By the Fokker-Planck equation:

$$\frac{d}{dt}p_j = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \partial_x \left(-\mu(x)\pi(x) + \partial_x D(x)\pi(x)\right) = \left[-\mu(x)\pi(x) + \partial_x D(x)\pi(x)\right]_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}}.$$

Let $J_{j-\frac{1}{2}}$ denote the flux through the $j-\frac{1}{2}$ boundary. Then the change in the probability that the system occupies the j^{th} cell is equivalent to the flux of probability through the boundaries:

$$\frac{d}{dt}p_j = J_{j+\frac{1}{2}} - J_{j-\frac{1}{2}} \tag{8.22}$$

where the flux is given by (see Equation (8.5)):

$$J_{j+\frac{1}{2}} = -\mu_{j+\frac{1}{2}}\pi_{j+\frac{1}{2}} + \partial_x D_{j+\frac{1}{2}}\pi_{j+\frac{1}{2}}.$$

In order to complete the discretization we need to approximate $\pi(x \pm \Delta x)$ and its derivatives given p. There is no unique way to make this approximation, however the following procedure is straightforward, sufficiently accurate, and interpretable.

The average value of $\pi(x)$ over the j^{th} cell is $\frac{1}{\Delta x}p_j$. Approximate $\pi(x)$ at the boundary between the j^{th} and $j+1^{st}$ cell with:

$$\pi_{j\pm\frac{1}{2}} \approx \frac{p_j + p_{j\pm1}}{2\Delta x}.$$

To approximate the derivative, use center differencing with $\pi(x_j) \approx p_j/\Delta x$:

$$\partial_x D_{j\pm\frac{1}{2}} \pi_{j\pm\frac{1}{2}} \approx \pm\frac{1}{\Delta x}\left(D_j \frac{p_j}{\Delta x} - D_{j\pm1}\frac{p_{j\pm1}}{\Delta x}\right) = \pm\frac{1}{\Delta x^2}\left(D_j p_j - D_{j\pm1}p_{j\pm1}\right).$$

Substitute these two approximations into the flux:

$$J_{j+\frac{1}{2}} = -\frac{1}{2\Delta x}\mu_{j+\frac{1}{2}}(p_j + p_{j+1}) + \frac{1}{\Delta x^2}\left(D_{j+1}p_{j+1} - D_j p_j\right).$$

Now the time derivative of p_j can be approximated as a linear combination of its neighbors:

$$\begin{aligned}\frac{d}{dt}p_j = &-\frac{1}{2\Delta x}[\mu_{j+\frac{1}{2}}(p_j + p_{j+1}) - \mu_{j-\frac{1}{2}}(p_j + p_{j-1})] \\ &+ \frac{1}{\Delta x^2}[D_{j+1}p_{j+1} - 2D_j p_j + D_{j-1}p_{j-1}].\end{aligned} \tag{8.23}$$

Notice that the first bracketed term is a finite difference approximation to $\partial_x \mu(x)p(x)$ and the second bracketed term is the standard central difference approximation to $\partial_x^2 D(x)p(x)$. Both of these approximations are second order accurate in Δx [76], so the discretization converges to the Fokker-Planck equation quadratically as Δx goes to zero.

To write the linear system in the standard master equation form rearrange the terms so that each p_j only appears once:

$$\frac{d}{dt}p_j = l_{j,j-1}p_{j-1} + l_{j,j}p_j + l_{j,j+1}p_{j+1} = Lp. \tag{8.24}$$

Where L is tridiagonal and:

$$\begin{bmatrix} l_{j,j-1} = \dfrac{1}{2\Delta x}\mu_{j-\frac{1}{2}} + \dfrac{1}{\Delta x^2}D_{j-1} \\ l_{j,j} = -\dfrac{1}{2\Delta x}(\mu_{j+\frac{1}{2}} - \mu_{j-\frac{1}{2}}) - \dfrac{2}{\Delta x^2}D_j \\ l_{j,j+1} = -\dfrac{1}{2\Delta x}\mu_{j+\frac{1}{2}} + \dfrac{1}{\Delta x^2}D_{j+1} \end{bmatrix}. \tag{8.25}$$

This accomplishes two of our goals. We have successfully found a second order accurate discretization scheme that produces a tridiagonal matrix L. To interpret L as a Laplacian it must conserve probability, so each column must sum to zero. Since L is tridiagonal the column sum only includes three terms:

$$\sum_{i,j} l_{i,j} = l_{j-1,j} + l_{j,j} + l_{j+1,j}.$$

To find $l_{j\pm1,j}$ note that $l_{j-1,j} = l_{(j-1),(j-1)+1}$ so can be recovered from Equation (8.25) by subtracting one from every index in $l_{j,j-1}$. Similarly $l_{j+1,j}$ can be found by adding one to every index in $l_{j,j-1}$. Then:

$$\begin{aligned} \sum_{i,j} l_{i,j} &= \left[-\frac{1}{2\Delta x}\mu_{j-\frac{1}{2}} + \frac{1}{\Delta x^2}D_j\right] - \left[\frac{1}{2\Delta x}(\mu_{j+\frac{1}{2}} - \mu_{j-\frac{1}{2}}) - \frac{2}{\Delta x^2}D_j\right] \\ &\quad + \left[\frac{1}{2\Delta x}\mu_{j+\frac{1}{2}} + \frac{1}{\Delta x^2}D_j\right] \\ &= 0. \end{aligned}$$

Therefore every column of L sums to zero, and the discretization conserves probability. This is an automatic virtue of working with a finite volume method [76].

To interpret L as a Laplacian all of its offdiagonal entries must be nonnegative, which requires $l_{j,j-1} \geq 0$ and $l_{j,j+1} \geq 0$. The off-diagonal entries of Equation (8.25) are not

automatically nonnegative, since advection may outweigh diffusion at the scale of the discretization. If the diffusion term is sufficiently small, or the drift term sufficiently large, then some of the off-diagonal elements of L may be negative. In that case L cannot be considered a Laplacian associated with a discrete-space continuous-time Markov process. Nonnegativity requires:

$$\frac{1}{\Delta x^2} D_{j\pm 1} \geq \frac{1}{2\Delta x} |\mu_{j\pm\frac{1}{2}}|.$$

Or:

$$\Delta x \leq 2 \frac{D_{j\pm 1}}{|\mu_{j\pm\frac{1}{2}}|}. \tag{8.26}$$

Thus, to interpret L as a Laplacian Δx must be sufficiently small. This can be accomplished either by setting the noise variance very large relative to the drift term, or by using a very fine discretization. In general the smaller the diffusion term and the larger the drift term the finer the discretization size Δx needs to be. Since advection and diffusion scale differently in the discretization size, $1/\Delta x$ and $1/\Delta x^2$ respectively, it is always possible to pick a discretization sufficiently small to guarantee accuracy.

Provided Δx is sufficiently small Equation (8.25) defines a second order accurate discretization of the Fokker-Planck equation such that L is tridiagonal, conserves probability, and has all non-negative off-diagonal entries. Then we can study the behavior of the master equation:

$$\frac{d}{dt} p = Lp$$

exactly as in Chapter 6 and Chapter 7. Then, as before, let:

$$l_{j,j\pm 1} = \rho_{j,j\pm 1} \exp\left(f_{j,j\pm 1}^{(L(\Delta x))}\right)$$

where the superscript is added to distinguish the edge flow from the forces.

Then:

$$\begin{aligned}\left[\frac{1}{2}(L-L^T)\right]_{j,j-1} &= \exp(\rho_{j,j-1})\sinh(f_{j,j-1}^{(L(\Delta x))}) \\ \left[\frac{1}{2}(L+L^T)\right]_{j,j-1} &= \exp(\rho_{j,j-1})\cosh(f_{j,j-1}^{(L(\Delta x))}).\end{aligned} \tag{8.27}$$

Substituting in for the forward and backward rates:

$$\begin{aligned}\frac{1}{2}(l_{j+1,j}-l_{j,j+1}) &= \frac{1}{2}\left(\frac{1}{2\Delta x}\mu_{j+1/2}+\frac{1}{2\Delta x}\mu_{j+1/2}+\frac{1}{\Delta x^2}D_j-\frac{1}{\Delta x^2}D_{j+1}\right) \\ &= \frac{1}{2\Delta x}\left(\mu_{j+1/2}-\frac{1}{\Delta x}(D_{j+1}-D_j)\right) \\ \frac{1}{2}(l_{j+1,j}+l_{j,j+1}) &= \frac{1}{2}\left(\frac{1}{2\Delta x}\mu_{j+1/2}-\frac{1}{2\Delta x}\mu_{j+1/2}+\frac{1}{\Delta x^2}D_j+\frac{1}{\Delta x^2}D_{j+1}\right) \\ &= \frac{1}{2\Delta x^2}(D_j+D_{j+1}).\end{aligned} \tag{8.28}$$

Then, to solve for f divide the first equation by the second:

$$\tanh\left(f^{(L(\Delta x))}\right) = \frac{\Delta x}{2}\frac{\mu_{j+1/2}-\frac{1}{\Delta x}(D_{j+1}-D_j)}{\frac{1}{2}(D_j+D_{j+1})}.$$

Then, expanding in Δx:

$$\begin{aligned}&\mu_{j+1/2} = \mu(x+\Delta x/2) \\ &\frac{1}{\Delta x}(D_{j+1}-D_j) = \partial_x D(x+\Delta x/2)+\mathcal{O}(\Delta x^2) \\ &\frac{1}{2}(D_j+D_{j+1}) = D(x+\Delta x/2)+\mathcal{O}(\Delta x^2).\end{aligned}$$

Therefore, the edge flow on the edge between x and $x+\Delta x/2$ is:

$$\tanh\left(f^{(L(\Delta x))}(x+\Delta x/2)\right) = \frac{\Delta x}{2}\frac{\mu(x+\Delta x/2)-\partial_x D(x+\Delta x/2)}{D(x+\Delta x/2)}+\mathcal{O}(\Delta x^3).$$

Taking the arc-tanh, and shifting the indexing by $\Delta x/2$:

$$f^{(L(\Delta x))}(x) = \text{atanh}\left(\frac{\Delta x}{2}\frac{\mu(x) - \partial_x D(x)}{D(x)} + \mathcal{O}(\Delta x^3)\right).$$

Notice that the term inside the parenthesis is exactly the forces we defined for the SDE scaled by $\Delta x/2$. To finish the analysis, Taylor expand the arctanh: $\text{atanh}(x) = x + \frac{1}{3}x^3 + \mathcal{O}(x^5)$. Then:

$$f^{(L(\Delta x))}(x) = \frac{\Delta x}{2}\frac{\mu(x) - \partial_x D(x)}{D(x)} + \mathcal{O}(\Delta x^3) = \frac{\Delta x}{2}f(x) + \mathcal{O}(\Delta x^3). \tag{8.29}$$

Thus, the edge flow converges to $\Delta x/2$ times the forces, exactly as in the hydrodynamic limit, and the approximation $\frac{2}{\Delta x}f^{(L(\Delta x))}(x) \simeq f(x)$ is $\mathcal{O}(\Delta x^2)$ accurate. Thus, at least in one dimension, the forces are consistent with the edge flow of a discrete-space continuous-time approximation to the SDE, and in the hydrodynamic limit of a random walk the edge flow converges to the forces.

8.4 Potentials in the Continuum

Given the forces $f(x) = D^{-1}(x)(\mu(x) - \nabla \cdot D(x))$ there are three natural potentials. These are presented here and compared.

8.4.1 The Helmholtz Potential in the Continuum

Let $f(x)$ be the forces. Then a function $\phi(x)$ is a Helmholtz potential $\phi(x)$ if it satisfies:

$$\begin{aligned} -\nabla\phi(x) + f_{\text{rot}}(x) &= f(x) \\ \nabla \cdot W(x) f_{\text{rot}}(x) &= 0. \end{aligned} \tag{8.30}$$

for some symmetric positive definite weight matrix $W(x)$. The weight matrix could be an identity, or could be the diffusion tensor $D(x)$. Combining the top and bottom equations yields the PDE:

$$\nabla \cdot W(x) \left(\nabla \phi(x) + f(x)\right) = 0. \tag{8.31}$$

The rotational component $f_{\text{rot}} = \nabla\phi(x) + f(x)$ is the residual left over when approximating $f(x)$ with the gradient of a potential function. A Helmholtz potential is a potential such that the residual is divergence free (when $W(x) = I$), or is divergence free in the weighted sense $\nabla \cdot W(x) f_{\text{rot}} = 0$. Both the Helmholtz potential and the quasipotential can be expressed in this way - the forces $f(x)$ are approximated with the gradient of a potential function, and a constraint is introduced on the residual that ensures that the residual is, in some sense, circulatory. Using the Helmholtz potential the residual must be divergence free so is incompressible. If $W(x) = I$ and the process is two or three dimensional then $f_{\text{rot}}(x)$ can be expressed as the curl of a vector potential, possibly with the addition of a harmonic component. Note that $W(x)$ plays essentially the same role that weights play in the weighted HHD (see Section 2.4)

Weights $W(x)$ could enter the decomposition after a coordinate transform. If $x = x(y)$, $y \in \mathbb{R}^n$ for some invertible change of coordinates and, in the original coordinate system $\nabla_x \cdot (\nabla_x \phi_x(x) + f_x(x)) = 0$, then in terms of y, $\nabla_y \cdot A(y)^{-\intercal} A(y)^{-1} \left(\nabla_y \phi_y(y) + f_y(y)\right) = 0$, where the subscripts on ∇ denote differentiation with respect to the associated coordinate, $A(y)$ is the Jacobian of $x(y)$, $\phi_y(y) = \phi_x(x(y))$ is the potential in the y coordinate system, and $f_y(y) = A(y) f_x(x(y))$ are the forces in the y coordinate system.

For example, consider the SDE $dX = \mu(X)dt + \sqrt{2}D^{1/2}(X)dW$ where $D^{1/2}(x)$ is symmetric. Then if there is an invertible change of coordinates $x = x(y)$ with Jacobian $A(y) = D^{1/2}(x(y))$ then $(x(y) - x(y_0)) \simeq A(y)(y - y_0) = D^{1/2}(y_0)(y - y_0)$, so $dX \simeq$

$D^{1/2}(x(y))dY$. Then $dY = D^{-1/2}(X)dX = D^{-1/2}(x(Y))\mu(x(Y)) + \sqrt{2}dW$, so $Y(t)$ is a stochastic process governed by an SDE with constant noise variance. Then if $\phi_x(x)$ is the solution to $\nabla_x \cdot D(x)(\nabla_x \phi_x(x) + f_x(x))$, then $\phi(x(y))$ is the solution to the unweighted equation $\nabla \cdot D^{-1/2}(x(y))D(x(y))D^{-1/2}(x(y))(\nabla \phi_y(y) + f_y(y))$ which equals $\nabla \cdot (\nabla \phi_y(y) + f_y(y))$. Therefore, if $W(x) = D(x)$, then the Helmholtz potential may be interpreted as the Helmholtz potential associated with the unweighted HHD after a change of coordinates into a coordinate system where the noise is isotropic and uniform. In that coordinate system the remainder f_{rot} is divergence free.

Note that [21] uses the unweighted convention $W(x) = I$ when decomposing the forces.

It is also important to note that the Helmholtz potential may not be unique for all fields $f(x)$. In $\mathbb{R}^3$ the potential is only unique if $f(x)$ vanishes faster than $1/||x||$ as $||x|| \to \infty$. See Section 2.2.1 or [8] for a review of conditions which guarantee uniqueness.

Continuum limit of HHD on a lattice

For a general sequence of reaction networks converging to an SDE the discrete HHD does not, in general, converge to the Helmholtz-Hodge decomposition of the forces. The discrete HHD does not necessarily converge to the Helmholtz-Hodge decomposition of the forces because the topology of the family of networks may not converge to an approximation of $\mathbb{R}^n$. Consider a square lattice with sides lengths Δx embedded in $\mathbb{R}^2$. Let $x \in \mathbb{R}^2$ denote the location of a vertex. Now suppose there are transitions between x and $x \pm \Delta x e_1$, $x \pm \Delta x e_2$, and we also add transitions from x to $x \pm \Delta x e_1 \pm \Delta x e_2$. Then the network consists of a square lattice plus diagonal edges between the corners of the lattice. In order for the discrete HHD to converge to the forces we must be able partition the edges so that each set of edges in the partitioning corresponds to a particular point in space, x, and each set

has two degrees of freedom corresponding to a basis of $\mathbb{R}^2$. There is no way to make this partitioning since there are too many edges in the network. If there were no diagonal edges in the network then we could associate the edges that transition between x and $x + \Delta x e_1$, $x + \Delta x e_2$ with node x, and would have a pair of edges for each vertex that correspond to the canonical basis in $\mathbb{R}^2$. Then the pair of edge flows associated with each node could converge to the pair of values in the force $f(x)$ associated with that point in space. Once the diagonal edge is added the edge flow has too many degrees of freedom, so we cannot establish a one-to-one mapping between the entries of the force at a given point in space, and a set of edges associated with a particular vertex. If the stoichiometry matrix associated with a reaction network is not invertible then there are too many possible reactions (edges) per node to establish a one-to-one mapping between the edge flow and the forces.

That said, if the sequence of networks is a sequence of square lattices formed by the Cartesian product of a refinement of the line with itself, then it is possible to show convergence of the discrete HHD to the HHD of the forces provided the transition rates are chosen so that the discrete-space process converges to the corresponding SDE.

Let $\mathcal{G}_0(\Delta x)$ be a refinement of $\mathbb{R}$ with equally spaced nodes separated by Δx. Then let $\mathcal{G}(\Delta x)$ be the Cartesian product $\square_{j=1}^{n} \mathcal{G}_0(\Delta x)$ which is the n dimensional hypercube lattice with sidelength Δx. Let x denote the coordinates of a particular vertex in $\mathbb{R}^n$. Let $\Omega(x)$ denote the hypercube centered at x with sidelengths Δx. Let $\partial\Omega(x)$ denote the boundary of $\Omega(x)$. Let $\partial\Omega_j^{\pm}(x) = \Omega(x) \cap \Omega(x \pm \Delta x e_j)$ denote the face shared between x and its neighbor $x \pm \Delta x e_j$. Let $\partial\Omega_{ji}^{\pm\pm} = \Omega(x) \cap \Omega(x \pm \Delta x e_j) \cap \Omega(x \pm \Delta x e_i) \cap \Omega(x \pm \Delta x e_j \pm \Delta x e_i)$ denote the shared boundary between the four vertices x, $x \pm \Delta x e_j$, $x \pm \Delta x e_i$, and $x \pm \Delta x e_j \pm \Delta x e_i$. In $\mathbb{R}^3$, $\partial\Omega_j^{\pm}$ is a face of the cube containing x, and $\partial\Omega_{ji}^{\pm\pm}$ is an edge of the cube. The notation introduced here is illustrated in Figure 8.2.

Let $p(x,t) = \iiint_{\Omega(x)} \pi(x,t)$ be the probability that $X(t) \in \Omega(x)$. Then the rate of

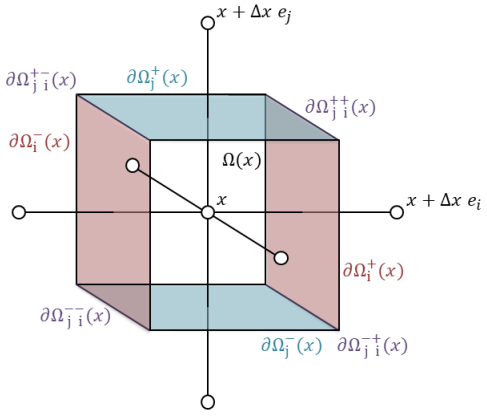

Figure 8.2: An example volume in a cubic lattice. The circles represent nodes. The cube surrounding the central node is $\Omega(x)$. The boundary faces associated with direction e_j and e_i are shaded blue and red respectively. The intersection of these faces are shown in purple.

change in $p(x,t)$ is given by the net flux through the boundary of $\Omega(x)$, which can be expressed as a sum over each pair of faces with the same orientation:

$$\frac{d}{dt}p(x,t) = \sum_{j=1}^{n} \iint_{\partial\Omega_j^+} J_j(y,t|\pi)dy - \iint_{\partial\Omega_j^-} J_j(y,t|\pi)dy.$$

Let $J_j^{\pm}(x) = \iint_{\partial\Omega_j^{\pm}} J_j(y,t|\pi)dy$ be the flux through the face $\partial\Omega_j^{\pm}$ (with outward normal set in the positive direction). Then substituting in for the flux we find:

$$\begin{aligned} J_j^{\pm}(x) = & \iint_{\partial\Omega_j^{\pm}(x)} [\mu_j(y) - \partial_{x_j} D_{jj}(y)]\pi(y,t)dy \\ & - \iint_{\partial\Omega_j^{\pm}(x)} D_{jj}(y)\partial_{x_j}\pi(y,t)dy \\ & - \sum_{i\neq j} \iint_{\partial\Omega_{ji}^{\pm+}(x)} D_{ij}(y)\pi(y,t)dy - \iint_{\partial\Omega_{ji}^{\pm-}(x)} D_{ij}(y)\pi(y,t)dy \end{aligned}$$

where the separation into terms associated with the diagonal entries of the diffusion tensor, and off-diagonal tensor, can be derived using integration by parts. The top line is associated

with advection, while the bottom two lines are associated with diffusion. The diffusive terms are split into two parts. The first part is diffusion through the face $\Omega_j^{\pm}(x)$, and the second is diffusion through the intersection of faces $\Omega_j^{\pm}(x)$ and $\Omega_i^{\pm}(x)$.

Next we need to approximate $\pi(y,t)$ and its partial derivatives on each face (and intersection of faces), given the probability $p(x,t)$ that $X(t)$ is in any of the volumes $\Omega(x)$. A natural extension of the approximation scheme used in one dimension is:

$$\pi(y,t) \simeq \frac{1}{\Delta x^n}\frac{(p(x)+p(x\pm\Delta x e_j))}{2} \text{ if } y \in \partial\Omega_j^{\pm}(x)$$
$$\pi(y,t) \simeq \frac{1}{\Delta x^n}\frac{(p(x)+p(x\pm\Delta x e_j)+p(x\pm\Delta x e_i)+p(x\pm\Delta x e_j\pm\Delta x e_i))}{4} \text{ if } y \in \partial\Omega_{ji}^{\pm\pm}(x)$$
$$\partial_{x_j}\pi(y,t) \simeq \frac{1}{\Delta x^n}\frac{\pm(p(x\pm\Delta x e_j)-p(x))}{\Delta x} \text{ if } y \in \partial\Omega_j^{\pm}(x).$$

Notice that the middle term, which is required to approximate the diffusion through the intersection of two faces, depends on the average of the probability of four neighboring volumes. Thus, if this approximation is substituted back into the expression for the flux through each face, the resulting finite volume method for approximating $\frac{d}{dt}p(x,t)$ will include flow of probability between the node representing x and $x+\Delta x(\pm e_j \pm e_i)$ for all i and j. Thus, the associated finite volume method would converge to a master equation on a graph that is not a square lattice in n dimensions. Instead it would converge to a master equation on the square lattice, with diagonal edges connecting nodes that differ in two coordinates. That is, instead of a graph that is given by a Cartesian product of the line segment with itself, the resulting graph would be the result of a strong product [64] of the line segment with itself. Our goal in this section is to show that edge flow on the network converges to the edge flow in the continuum *and* the operators of the discrete HHD converge to the corresponding differential operators in the continuum, so the entire decomposition converges. Convergence of this kind is only possible if the

underlying sequence of networks are square lattices, and do not include extra diagonal edges. Therefore, from now on we will assume that the diffusion tensor $D(x)$ is diagonal for all x. Note that, if the SDE arises from a Langevin approximation to a reaction network then the diffusion tensor is automatically diagonal if all the reactions only change one state variable at a time, as in a birth-death process. Alternatively, if $D(x)$ is diagonalizable, and can be diagonalized by an orthonormal set of eigenvectors that change smoothly in x, then it may be possible to construct a sequence of networks that converge to a rotated square lattice for vanishing neighborhoods about each point, such that the lattices are oriented to diagonalize $D(x)$.

If $D(x)$ is diagonal, then, substituting the finite volume approximation in for $\pi(x,t)$ in the expression for $J_j^{\pm}(x)$ gives:

$$J_j^{\pm}(x) \simeq \frac{1}{\Delta x^n}\left[\iint\limits_{\partial\Omega_j^{\pm}(x)} \mu_j(y) - \partial_{x_j} D_{jj}(y) dy\right] \frac{p(x \pm \Delta x e_j) + p(x)}{2} - \frac{1}{\Delta x^n}\left[\iint\limits_{\partial\Omega_j^{\pm}(x)} D_{jj}(y) dy\right] \frac{\pm(p(x \pm \Delta x e_j) - p(x))}{\Delta x}.$$

Let $g(x)$ be an arbitrary real scalar valued function. Then let $\overline{g}(A)$ be the average value of $g(x)$ over the set A. Then:

$$J_j^{\pm}(x) \simeq \frac{1}{\Delta x}\left[\overline{\mu_j}(\partial\Omega_j^{\pm}(x)) - \overline{\partial_{x_j} D_{jj}}(\partial\Omega_j^{\pm}(x))\right] \frac{p(x \pm \Delta x e_j) + p(x)}{2} - \frac{1}{\Delta x}\left[\overline{D_{jj}}(\partial\Omega_j^{\pm}(x))\right] \frac{\pm(p(x \pm \Delta x e_j) - p(x))}{\Delta x}.$$

Then, the forward rate of transition from x to $x + \Delta x e_j$, and corresponding backwards

rates are:

$$l_j^+(x) = \frac{1}{2\Delta x}\left[\overline{\mu_j}(\partial\Omega_j^\pm(x)) - \overline{\partial_{x_j} D_{jj}}(\partial\Omega_j^\pm(x))\right] + \frac{1}{\Delta x^2}\overline{D_{jj}}(\partial\Omega_j^\pm(x))$$

$$l_j^-(x+\Delta x e_j) = -\frac{1}{2\Delta x}\left[\overline{\mu_j}(\partial\Omega_j^\pm(x)) - \overline{\partial_{x_j} D_{jj}}(\partial\Omega_j^\pm(x))\right] + \frac{1}{\Delta x^2}\overline{D_{jj}}(\partial\Omega_j^\pm(x))$$

Note that these two rates involve averages over the same face, since they both depend on the flux across the face separating $\Omega(x)$ and $\Omega(x+\Delta x e_j)$.

Now let $f_j^{(L(\Delta x))}\left(x+\frac{\Delta x}{2}e_j\right)$ be the edge flow on the edge from x to $x+\Delta x e_j$. Then, repeating the same technique we used for the one-dimensional case:

$$f_j^{(L(\Delta x))}\left(x+\frac{\Delta x}{2}e_j\right) = \tanh^{-1}\left(\frac{\Delta x}{2}\frac{\overline{\mu_j}(\partial\Omega_j^\pm(x)) - \overline{\partial_{x_j} D_{jj}}(\partial\Omega_j^\pm(x))}{\overline{D_{jj}}(\partial\Omega_j^\pm(x)}\right).$$

So, in the limit Δx goes to zero, the edge flow on each edge in the discrete approximation to the SDE is:

$$f_j^{(L(\Delta x))}\left(x+\frac{\Delta x}{2}e_j\right) = \frac{\Delta x}{2}\left[\overline{D}(\partial\Omega_j^\pm(x))^{-1}\left(\overline{\mu}(\partial\Omega_j^\pm(x)) - \overline{\nabla\cdot D}(\partial\Omega_j^\pm(x))\right]_j + \mathcal{O}(\Delta x^3). \tag{8.32}$$

Now, in the continuum the forces in the direction e_j at x are $D^{-1}(x)(\mu(x)-\nabla\cdot D(x))$. Then $\left[\overline{D}(\partial\Omega_j^\pm(x))^{-1}\left(\overline{\mu}(\partial\Omega_j^\pm(x)) - \overline{\nabla\cdot D}(\partial\Omega_j^\pm(x))\right]_j$ is a $\mathcal{O}(\Delta x^2)$ accurate approximation to $f_j\left(x+\frac{\Delta x}{2}e_j\right)$ so:

$$f_j^{(L(\Delta x))}\left(x+\frac{\Delta x}{2}e_j\right) = \frac{\Delta x}{2}f_j\left(x+\frac{\Delta x}{2}e_j\right) + \mathcal{O}(\Delta x^3). \tag{8.33}$$

Equation (8.33) establishes that the forces on each edge in the discrete lattice approximation to an SDE in $\mathbb{R}^n$ with a diagonal diffusion tensor converge to the forces scaled by

Δx.

The corresponding discrete HHD reads:

$$-G\phi^{(L(\Delta x))} + C^T\theta^{(L(\Delta x))} = f^{(L(\Delta x))}.$$

Divide both sides by Δx. Then, on each edge, we have:

$$\left[-\frac{1}{\Delta x}G\phi^{(L(\Delta x))} + \frac{1}{\Delta x}C^T\theta^{(L(\Delta x))}\right]_j \left(x + \frac{\Delta x}{2}e_j\right) = \frac{1}{2}f_j\left(x + \frac{\Delta x}{2}e_j\right) + \mathcal{O}(\Delta x^2).$$

In this form the left hand side consists of a pair of discrete approximations to differential operators, and the right hand side converges to the forces defined by Equation (8.16). In order for the discrete potentials to converge to the potentials defined by the continuous HHD we first show that the operators converge. Then we show that the discrete weighted Poisson equation associated with a weighted version of the HHD converges to the weighted Poisson equation used to define the Helmholtz potential in the continuum. This establishes the conceptual equivalence of the potentials associated with the discrete HHD and the Helmholtz potential. To conclude we show that, in the unweighted case, the discrete scalar potential converges to the continuous potential using the spectral approach to solving the discrete Poisson equations developed in Section 3.3.3.

Convergence of the gradient is trivial. Let u be a differentiable function on $\mathbb{R}^n$, and sample u at each node x in the graph. Then:

$$\left[\frac{1}{\Delta x}Gu\right]_j \left(x + \frac{\Delta x}{2}e_j\right) = \frac{u(x + \Delta x e_j) - u(x)}{\Delta x} = \partial_{x_j}u\left(x + \frac{\Delta x}{2}e_j\right) + \mathcal{O}(\Delta x^2) \tag{8.34}$$

Therefore the discrete gradient applied to u converges, on each edge, to the corresponding partial derivative of u evaluated at the center of the edge.

Illustrating convergence for the curl is, not surprisingly, a little more difficult. To help organize the calculation subdivide the loops in the square space of the lattice into sets of loops with a specific orientation. Consider a node x in the lattice. From each node we can uniquely specify a face $\mathcal{C}^{ij}(x)$ of the lattice where $\mathcal{C}^{ij}(x)$ is the face contained between the nodes $[x, x+e_j, x+e_i, x+e_j+e_i]$. Notice that $\mathcal{C}^{ij}(x)$ is the same face as $\mathcal{C}^{ji}(x)$ and $\mathcal{C}^{ii}(x)$ does not correspond to a face. Therefore there are $d(d-1)/2$ faces per node. Each face corresponds to a particular plane of rotation. Define a vector valued function $\Theta(x) = [\theta^{12}(x), \theta^{13}(x), ...\theta^{d-1,d}(x)]$. The function $\Theta(x)$ has $d(d-1)/2$ outputs for every point x. It will play the same role as the vector potential in electromagnetism. In terms of our standard decomposition this corresponds to defining $d(d-1)/2$ sets of loops $(1,2),(1,3),...(d-1,d)$ that correspond to specific orientations of rotation. In essence these definitions partition the set of loops in the square space of the lattice into subsets associated with each node. Recall that the set of all loops in the square space is too large to be a cycle basis. However, as noted in Section 2.4 we can work with an extended cycle basis without changing the scalar potential, conservative component, or rotational component of the decomposition.

Since the adjoint curl maps from the space of loops to the space of edges this partitioning of the loops into sets with specific orientations corresponds to partitioning the adjoint curl C^T into a set of operators $C^{ij\,T}$ where $C^{ij\,T}$ maps from the loops (i,j).

With this modified notation:

$$C^T\Theta(x) = \sum_{i<j} C^{ij\,T}\theta^{ij}(x).$$

The k^{th} entry of this product evaluated at x corresponds to the edge from x to $x+\Delta x e_k$.

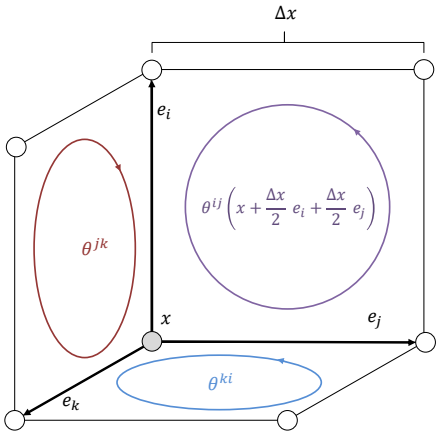

Figure 8.3: The three different loop classes (ij,jk and ki) neighboring a given node x in a cubic lattice.

Evaluating the product (scaled by Δx) gives:

$$\left[\frac{C^{ij^T}}{\Delta x}\theta^{ij}(x)\right]_k = \begin{cases} 0 & \text{if } k \neq i \text{ or } j \\ \dfrac{\theta^{ij}(x+\Delta x e_j/2)-\theta^{ij}(x-\Delta x e_j/2)}{\Delta x} & \text{if } k=i \\ \dfrac{\theta^{ij}(x-\Delta x e_i/2)+\theta^{ij}(x-\Delta x e_i/2)}{\Delta x} & \text{if } k=j \end{cases}$$

Therefore, in the limit Δx goes to zero:

$$\lim_{\Delta x\to 0}\left[\frac{C^{ij^T}}{\Delta x}\theta^{ij}(x)\right]_k = \begin{cases} 0 & \text{if } k \neq i \text{ or } j \\ \partial_{x_j}\theta^{ij}(x) & \text{if } k=i \\ -\partial_{x_i}\theta^{ij}(x) & \text{if } k=j \end{cases}$$

Now, define a set of $d \times d$ rotation matrices R^{ij} which take j to i and i to $-j$, and send

all other entries to zero[5] then:

$$\lim_{\Delta x \to 0} \left[\frac{C^{ij^T}}{\Delta x} \theta^{ij}(x) \right]_k = R^{ij} \nabla \theta^{ij}(x)$$

Therefore, in three dimensions:

$$\lim_{\Delta x \to 0} \left[\frac{C^T}{\Delta x} \Theta(x) \right]_k = \sum_{i<j} R^{ij} \nabla \theta^{ij}(x) = \nabla \times \Theta(x). \tag{8.35}$$

Thus, if the curl is defined using all the loops in the square space then each entry of the curl of Θ is the same as the curl of Θ projected onto a hyperplane spanned by e_i and e_j for some i and j. This establishes that the discrete curl converges to a differential operator analogous to the curl in the continuum.

In general we can solve for the scalar potential without ever solving for a vector potential. Our objective is to show the conceptual equivalence of the Helmholtz potential defined by Equation (8.31) to the scalar potential arising from a weighted HHD on a discrete approximation to the SDE, not to show equivalence of both potentials. Therefore we will now focus exclusively on the scalar potential.

The scalar potential associated with a weighted Helmholtz-Hodge Decomposition of the edge flow is the solution to a discrete weighted Poisson equation:

$$\frac{1}{\Delta x^2} G^{(L(\Delta x))^\intercal} W G^{(L(\Delta x))} \phi^{(L(\Delta x))} = -\frac{1}{\Delta x} G^{(L(\Delta x))^\intercal} f^{(L(\Delta x))}.$$

[5] For example, if $d = 3$ then:

$$R^{12} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The right hand side converges to the divergence of the forces since:

$$\left[-\frac{1}{\Delta x}G^{(L(\Delta x))^\intercal}f^{(L(\Delta x))}\right](x) = \sum_{j=1}^{n}\frac{1}{\Delta x}\left[f_j\left(x+\frac{\Delta x}{2}e_j\right) - f_j\left(x-\frac{\Delta x}{2}e_j\right)\right] \to \nabla\cdot f(x).$$

Let $u(x)$ be a twice differentiable scalar valued function, which is sampled at the nodes of the lattice. Then:

$$\begin{aligned}
&\left[\frac{1}{\Delta x^2}G^{(L(\Delta x))^\intercal}WG^{(L(\Delta x))}u\right](x)\\
&\quad= \frac{1}{\Delta x}\sum_{j=1}^{n} w_j\left(x+\frac{\Delta x}{2}e_j\right)\frac{u(x+\Delta x e_j)-u(x)}{\Delta x} - w_j\left(x-\frac{\Delta x}{2}e_j\right)\frac{u(x)-u(x-\Delta x e_j)}{\Delta x}\\
&\quad\to \nabla\cdot W(x)\nabla u(x).
\end{aligned}$$

Therefore the operator on the left hand side of the discrete Poisson equation converges to $\nabla\cdot W\nabla$ as $\Delta x\to 0$, which is the same operator used to define the Helmholtz potential, and the right hand side of both equations converge. It remains to show that the potentials converge.

Here we show that the potentials converge in the special case when no weights are introduced, and $X(t)$ can reach any point inside a rectangular domain, but cannot leave the domain. That is, $X(t)$ is restricted to taking on values inside the Cartesian product of a sequence of intervals. For simplicity we will assume that the domain is the unit cube in n-dimensions. The following results could be generalized to arbitrary rectangular domains.

Then, the discrete approximations consist of a sequence of lattices formed by the Cartesian product of a line segment with itself n times. The segment has $1/\Delta x+1$ vertices, and each edge is length Δx. If the HHD is unweighted then the spectral approach to solving the discrete Poisson equation can be used. Our goal is to show that this spectral approach gives a convergent solution to the spectral approach in the continuum.

The spectral approach on the lattice depended on expanding the divergence of the edge flow onto the eigenbasis of the Laplacian. The eigenvectors of the Laplacian of the lattice are outerproducts of the eigenvectors of the Laplacian in one-dimension, and the eigenvalues are the sums of the eigenvalues of the Laplacian in one-dimension. In one-dimension the eigenvalues and vectors are:

$$\lambda_j = 4\sin^2\left(\frac{\pi(j-1)}{2V}\right) = \left(\frac{\pi(j-1)}{V}\right)^2 + \mathcal{O}(V^{-4})$$
$$v_{ij} = \left\{\begin{array}{l} V^{-1/2} \text{ if } j = 1 \\ \sqrt{\frac{2}{V}}\cos\left(\frac{\pi(j-1)}{V}\left(i-\frac{1}{2}\right)\right) \text{ else} \end{array}\right\}$$

where $V = \frac{1}{\Delta x} + 1$ is the number of vertices, and diverges as Δx becomes small. To convert to the continuous Poisson equation we divide both sides of the discrete Poisson equation by Δx^2. Then the Laplacian is divided by Δx^2 and converges to a differential operator. It follows that the eigenvalues of the scaled equations are:

$$\lambda_j = (\pi(j-1))^2 + \mathcal{O}(\Delta x^2).$$

Then let:

$$\omega(j) = \pi(j-1)$$

and let $x(i) = \Delta x(i-1)$. Then the eigenvalues and eigenvectors are of the form:

$$\lambda_j \simeq \omega(j)^2 + \mathcal{O}(\Delta x^2)$$
$$v_{ij} \simeq \left\{\begin{array}{l} (\Delta x)^{1/2} \text{ if } j = 1 \\ \sqrt{2\Delta x}\cos\left(\omega(j)(x(i) + \Delta x/2)\right) \text{ else} \end{array}\right\}.$$

That is, the eigenvectors are trigonometric functions in x, and the corresponding eigenvalues converge to the frequencies squared.

The eigenvectors and eigenvalues of the continuous Laplacian have the same structure since, in one dimension $-\partial_x^2 \exp(i\omega x) = \omega^2 \exp(i\omega x)$ and in higher dimension:

$$-\nabla \cdot \nabla \prod_{j=1}^{n} \exp(i\omega_j x_j) = -\sum_{j=1}^{n} \partial_{x_j}^2 \prod_{j=1}^{n} \exp(i\omega_j x_j) = \left(\sum_{j=1}^{n} \omega_j^2 \right) \prod_{j=1}^{n} \exp(i\omega_j x_j).$$

Moreover, if the domain has reflecting boundaries then the normal derivative of $\pi(x,t)$ must be zero at the boundaries, so the eigenfunctions are limited to cosines with frequencies such that the boundaries occur at extrema of the eigenfunctions. These are given by frequencies $\pi(j-1)$.

Thus, in the continuum the eigenvectors of the Laplacian are any product of cosine functions in each coordinate, with frequency such that the normal derivative at the boundary is zero, and with eigenvalues equal to the sum of squares of the frequencies. In the discrete case the eigenvectors are outerproducts of cosine functions with frequencies such that the normal derivative of the eigenfunction is vanishing at the boundary, and with eigenvalues that converge to the sums of the frequencies of squared. The convergence is faster at low frequencies than at high frequencies. Convergence of the eigenvectors ensures that the inner products used to move into and out of the eigenbasis converge, and convergence of the eigenvalues ensures convergence of the solutions in the eigenbasis (frequency space). Note that the eigenvalues converge fastest at low frequencies, so convergence requires forces whose divergence is sufficiently smooth. For more details on convergence see [73].

In summary, if the SDE has a diagonal diffusion tensor, as in a birth-death process, then it can be approximated by a family of lattices, whose edge-flow, gradient operator, and Laplacian converge to the forces, gradient, and weighted Laplacian. This establishes

the conceptual equivalence of the Helmholtz potential defined by Equation (8.31), and the scalar potential which solves a weighted HHD.

8.4.2 The Quasipotential

So far we have focused on extending the HHD based potentials theory we developed for networks into the continuum. The natural extension of the discrete scalar potential defined using the HHD is the Helmholtz potential defined by Equation (8.31). For SDEs an alternative potential exists, and is widely used - the quasipotential.

In the previous chapter we showed that the limiting behavior of the steady state distribution of a Markov process was governed by different potentials if the process was dominated by drift or if the process was dominated by diffusion. If the process was dominated by diffusion than an HHD type potential governed the steady state and steady state fluxes. In contrast, if the forcing was strong then advection dominated drift, and as a result a quasipotential framework was needed. In Section 2.3.2 we showed that differences in the scalar potential associated with an HHD is equivalent to the average work required to move between nodes, where the average is evaluated over an ensemble of paths. In contrast, when the forcing was strong, then a quasipotential was introduced, where the difference in quasipotential at pairs of nodes was equal to the work to move between those nodes over an optimal path. When a near deterministic limit was used then the optimal paths were the most likely sequence of nodes.

As in the discrete case, the quasipotential in the continuum is defined by evaluating the work over an optimal set of paths [23]. Quasipotentials are widely used to study multistable systems in small noise limits and large deviations. Applications of the quasipotential include modeling insect outbreaks, epidemics, extinction, invasion, and cell development (cf. [23, 277]).

Consider a path $y(t)$ on the time interval $[0, T]$. The probability of observing the path y can always be written:

$$\pi(y) \propto \exp\left(-S(y)\right) \tag{8.36}$$

where S is an action functional that maps trajectories to real numbers [24].

To approximate the action functional consider a sequence of evenly spaced sample times at intervals Δt. Let $y_j = y(t_j)$ where $t_j = j\Delta t$. Then the probability of observing a sample trajectory y can be approximated by computing the probability of sampling y using a numerical approximation scheme. The simplest such scheme is the Euler-Maruyama scheme, where $Y_{j+1} = Y_j + \mu_j(Y_j)\Delta t + B(Y_j)\Delta W$. Then:

$$\pi(y|\Delta t) = \prod_j \frac{1}{\sqrt{(2\pi)^n |D(y_j)|\Delta t^n}} \exp\left(-\frac{1}{2}||(y_{j+1} - y_j) - \mu(y_j)\Delta t||^2_{D^{-1}(y_j)\Delta t}\right).$$

Therefore:

$$\pi(y|\Delta t) \propto \exp(-\sum_j \frac{1}{2}||(y_{j+1} - y_j) + \mu(y_j)\Delta t||^2_{D^{-1}(y_j)\Delta t} - \frac{1}{2}\log(|D(y_j)|)).$$

Notice that $||(y_{j+1} - y_j) + \mu(y_j)\Delta t||^2_{D^{-1}(y_j)\Delta t}$ equals $||\frac{1}{\Delta t}(y_{j+1} - y_j) + \mu(y_j)||^2_{D^{-1}(y_j)}\Delta t$. Then, the discrete time action is:

$$S(y|\Delta t) = \sum_j \left(\frac{1}{2}||\frac{1}{\Delta t}(y_{j+1} - y_j) - \mu(y_j)||^2_{D^{-1}(y_j)}\Delta t + \frac{1}{2}\log(|D(y_j)|)\right).$$

Then the probability of any sample trajectory y is proportional to $\exp(-S(y|\Delta t))$. A continuous time action functional can then be defined by taking the limit as Δt goes to zero. Care has to be taken when defining this limit since sample trajectories of an SDE are not differentiable, and the probability of trajectories must be replaced with a probability

density.

Now suppose we consider the set of all paths $\mathcal{Y}$ starting from x_0 at time 0 and ending at x at time T. For any x there exists some path, or set of paths, that minimizes the action in the limit as the diffusion tensor vanishes. Let x_0 be a stable equilibrium of the deterministic skeleton. The infimum of the action over all possible paths from x_0 to x in the limit as the noise vanishes is the Friedlin-Wentzell quasipotential associated with the equilibrium x_0 [24, 25].

The optimal trajectories can be approximated using a variational approach. Broadly speaking, a Hamiltonian approach can be used to derive a pair of Euler-Lagrange equations that govern the motion of optimal trajectories. The optimal trajectories can be simulated by solving the system of ODEs defined by the Euler-Lagrange equations. These require introducing a “momentum" term. Fixing an initial position and momentum fixes an optimal trajectory. Note that this does not specify the endpoint of the optimal trajectory, so typically many trajectories using different initial momentum are generated to approximate the quasipotential surface. Using this approach it can be shown that, if the SDE obeys detailed balance, then the optimal trajectories are time-reversed trajectories of the deterministic process.

The quasipotential surface can also be defined as the solution to a PDE. The appropriate PDE is the Hamilton-Jacobi equation, which can be derived using a WKB expansion of the steady state [23, 24]. We will rely on this approach here since it is simpler, does not require as advanced machinery, and enables a more direct comparison to the Helmholtz potential.

The Wentzell-Kramers-Brillouin approximation, or WKB, is an approximation scheme for solving linear partial differential equations. The technique revolves around asymptotic expansion of a function in some parameter, and is an example of multiple-scale anaylsis. The relevant parameter here is the noise intensity σ. Replace $D(x)$ with $\sigma^2 D(x)$ so that

the noise intensity is controlled by a single parameter. For example, in a system size expansion σ^2 would scale in one over the system size. By varying σ it is possible to solve for approximate solutions to the Fokker-Planck equation. In particular the WKB can be applied to approximate the stationary distribution.

The main virtue of the WKB is that it can be used to accurately estimate the likelihood of rare events. It can be applied to estimate mean first passage times and to construct quasipotentials [24]. Example applications in ecology include estimating mean times to extinction [278].

The WKB is defined for linear differential equations whose highest derivative is scaled by some small parameter. The Fokker-Planck equation is a second order linear equation of this type, whose highest derivative is scaled by the noise intensity σ^2.

Given an n^{th} order equation of the form:

$$\epsilon \frac{d^n}{dx^n} f(x) + \alpha_{n-1}(x) \frac{d^{n-1}}{dx^{n-1}} f(x) + ... + \alpha_0 f(x) = 0$$

the WKB approximation proceeds by solving for a function $S(x|\epsilon)$ such that:

$$f(x) \propto \exp\left(-S(x|\epsilon)\right) \tag{8.37}$$

where $S(x|\epsilon)$ can be expanded using a Laurent series in ϵ [279]:

$$S(x|\epsilon) = \frac{1}{\epsilon} \sum_{m=0}^{\infty} S_m(x) \epsilon^m. \tag{8.38}$$

To solve for each term in the expansion substitute $S(x|\epsilon)$ into the differential equation, then take the limit as ϵ goes to zero. This gives a lower order equation exclusively in terms of $S_0(x)$. To compute higher order terms S_m, equate the appropriate orders in ϵ [279].

First let $q(x)$ denote the steady state distribution. Then write:

$$q(x) \propto \exp(-S(x|\sigma)) \tag{8.39}$$

where $S(x) = -\log(q(x))$ is an effective potential.

The first term in the WKB expansion is the small noise limit of the effective potential:

$$S_0(x) = \lim_{\sigma \to 0} \sigma^2 \log\left(q(x)\right) = \lim_{\sigma \to 0} S(x|\sigma). \tag{8.40}$$

Here we start by applying WKB to the Fokker-Planck equation in one dimension. The results mirror our previous analysis using integration by parts. The one-dimensional case is used primarily as an example to introduce the technique.

The steady state Fokker-Planck equation in one-dimension is:

$$-\frac{d}{dx}(\mu(x)q(x)) + \sigma^2 \frac{d^2}{dx^2}(D(x)q(x)) = 0.$$

To simplify, separate the derivatives using the product rule:

$$\begin{aligned}
&\sigma^2 \left(\left[\frac{d^2}{dx^2} D(x) \right] q(x) + 2 \frac{d}{dx} D(x) \frac{d}{dx} q(x) + D(x) \left[\frac{d^2}{dx^2} q(x) \right] \right) \\
&- \mu(x) \frac{d}{dx} q(x) - \left[\frac{d}{dx} \mu(x) \right] q(x) = 0.
\end{aligned}$$

Collecting terms, and dividing through by $D(x)$ yields:

$$\sigma^2 \frac{d^2}{dx^2} q(x) + \left[\frac{2\sigma^2 \frac{d}{dx} D(x) - \mu(x)}{D(x)} \right] \frac{d}{dx} q(x) + \frac{\sigma^2 \frac{d^2}{dx^2} D(x) - \frac{d}{dx} \mu(x)}{D(x)} q(x) = 0.$$

Then make the substitution: $q(x) = \exp(-S(x|\sigma))$. The derivatives of the exponential

are:

$$\frac{d}{dx}q(x) = -\left(\frac{d}{dx}S(x|\sigma)\right)\exp\left(-S(x|\sigma)\right)$$
$$\frac{d^2}{dx^2}q(x) = \left(-\frac{d^2}{dx^2}S(x|\sigma) + \left(\frac{d}{dx}S(x|\sigma)\right)^2\right)\exp\left(-S(x|\sigma)\right)$$

Plugging in, and dividing through by $-q(x) = -\exp(-S(x|\sigma))$ gives:

$$\sigma^2\left(\frac{d^2}{dx^2}S(x|\sigma) - \left(\frac{d}{dx}S(x|\sigma)\right)^2\right) - \left[\frac{2\sigma^2\frac{d}{dx}D(x) - \mu(x)}{D(x)}\right]\frac{d}{dx}S(x|\sigma) - \frac{\sigma^2\frac{d^2}{dx^2}D(x) = \frac{d}{dx}\mu(x)}{D(x)}.$$

Now expand $S(x|\sigma)$ with a Laurent series:

$$S(x|\sigma) = \frac{1}{\sigma^2}\sum_{m=0}^{\infty}S_m(x)\sigma^{2m} = \frac{1}{\sigma^2}S_0(x) + \sum_{m=1}^{\infty}S_m(x)\sigma^{2(m-1)}.$$

To recover the lowest order term, substitute the Laurent expansion into each term in the differential equation one at a time, and take the limit as σ goes to zero. The highest order term in the differential equation converges to:

$$\lim_{\sigma\to 0}\sigma^2\left(\frac{1}{\sigma^2}\sum_{m=0}^{\infty}\frac{d^2}{dx^2}S_m(x)\sigma^{2m} - \left(\frac{1}{\sigma^2}\sum_{m=0}^{\infty}\frac{d}{dx}S_m(x)\sigma^{2m}\right)^2\right)$$
$$= \lim_{\sigma\to 0} -\frac{1}{\sigma^2}\left(\frac{d}{dx}S_0(x)\right)^2 + \mathcal{O}(1).$$

The next highest order term converges to:

$$\lim_{\sigma\to 0}\left[\frac{2\sigma^2\frac{d}{dx}D(x) - \mu(x)}{D(x)}\right]\frac{1}{\sigma^2}\sum_{m=0}^{\infty}\frac{d}{dx}S_m(x)\sigma^{2m} = \lim_{\sigma\to 0}\frac{1}{\sigma^2}\left[\frac{-\mu(x)}{D(x)}\right]\frac{d}{dx}S_0(x) + \mathcal{O}(1).$$

The final term remains $\mathcal{O}(1)$ as σ goes to zero so:

$$\lim_{\sigma\to 0}\frac{-1}{\sigma^2}\left(\frac{d}{dx}S_0(x)\right)^2 + \frac{1}{\sigma^2}\left[\frac{-\mu(x)}{D(x)}\right]\frac{d}{dx}S_0(x) = 0.$$

Multiplying through by σ^2 gives:

$$\left(\frac{d}{dx}S_0(x)\right)^2 + \left[\frac{\mu(x)}{D(x)}\right]\frac{d}{dx}S_0(x) = 0.$$

This is a quadratic equation in $\frac{d}{dx}S_0$ therefore there are two possible solutions. Either:

$$\frac{d}{dx}S_0(x) = 0 \text{ or } \frac{d}{dx}S_0(x) = \frac{-\mu(x)}{D(x)}.$$

Therefore, after integration, $S_0(x)$ is either equal to some constant, or:

$$S_0(x) = \int_{x_0}^{x} \frac{-\mu(y)}{D(y)}dy.$$

This result is completely consistent with the small noise limit of the effective potential found via integration by parts (see Equation (8.17)).

Suppose now that $x \in \mathbb{R}^n$. Then, from Fokker-Planck, the equilibrium distribution obeys:

$$-\nabla \cdot (\mu(x)q(x)) + \sigma^2 \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}(x)q(x)) = 0.$$

To simplify the notation let ∂_i represent the partial derivative with respect to x_i and ∂_{ij}^2 represent the second partial derivative with respect to x_i and x_j.

In order to solve for S_0 rearrange the stationary Fokker-Planck equation. First, by the divergence product rule:

$$-\nabla \cdot (\mu(x)q(x)) = -[(\nabla \cdot \mu(x))q(x) + \mu(x) \cdot \nabla q(x)].$$

To expand the diffusion term use the standard product rule:

$$\sum_{ij} \partial_{ij}^2 (D(x)q(x)) = \sum_{ij} \left(\partial_{ij}^2 D_{ij}(x)\right) q(x) + 2\left(\partial_i D_{ij}(x)\right)\left(\partial_j q(x)\right) + D_{ij}(x)\partial_{ij}^2 q(x).$$

Now, letting $q(x) = \exp\left(-S(x|\sigma)\right)$:

$$\begin{aligned}
\partial_i q(x) &= -\left[\partial_i S(x|\sigma)\right] q(x) \\
\partial_{ij}^2 q(x) &= \left[-\partial_{ij}^2 S(x|\sigma) + \partial_i S(x|\sigma)\partial_j S(x|\sigma)\right] q(x).
\end{aligned}$$

All terms are scaled by $q(x)$ so we can divide it out from the full expression. To find S_0 consider the limit that σ goes to zero. Then the terms $\mathcal{O}(\sigma^{-2})$ dominate the equation. This leaves:

$$\nabla S_0(x) \cdot \mu(x) + (\nabla S_0(x))^T D(x)(\nabla S_0(x)) = 0.$$

or:

$$\nabla S_0(x) \cdot (D(x)\nabla S_0(x) + \mu(x)) = 0.$$

Then, provided the diffusion tensor is invertible:

$$\nabla S_0(x) \cdot D(x)\left(\nabla S_0(x) + D^{-1}(x)\mu(x)\right). \tag{8.41}$$

Equation (8.41) is a static Hamilton-Jacobi equation [23]. Hamilton-Jacobi equations play a fundamental role in classical mechanics, in particular, for the trajectory of particles that conserve energy. The Hamilton-Jacobi equation has three solutions. First, the parenthetical term could be zero:

$$\nabla S_0(x) = -D^{-1}(x)\mu(x) \tag{8.42}$$

which would require that the vector field $D^{-1}(x)\mu(x)$ is conservative, and in which case

$S_0(x)$ would be the associated potential. Second, the second term could be zero:

$$\nabla S_0(x) = 0$$

which would require that the potential is constant. Or, third, the two terms could be orthogonal. This indeterminacy allows for multiple solutions of the first type joined along surfaces where $\nabla S_0(x) = 0$, or $\nabla S_0(x)$ is perpendicular to $\left[-\mu(x) + \frac{1}{2}D(x)(\nabla S_0(x))\right]$. In general we will focus on solutions of the first type or third type inside domains where the latter solutions are not possible. To construct a general solution we can combine these domains, making sure to match S_0 at the boundaries appropriately. This is the procedure suggested in [23]. Numerical methods are provided in [280].

The first solution generalizes the one dimensional result $\partial S_0(x) = -\frac{\mu(x)}{D(x)}$ and is generally the most intuitive. If $D(x)^{-1}\mu(x)$ is a gradient system then there exists an $S_0(x)$ such that $\nabla S_0 = -D(x)^{-1}\mu(x)$. This is the standard relationship between a vector field and a potential.

If $D(x)^{-1}\mu(x)$ is not a gradient system then we need to find an S_0 such that the gradient of $S_0(x)$ is perpendicular to the difference between $-D(x)^{-1}\mu(x)$ and $\nabla S_0(x)$.

Let $\mu_{circ} = \nabla S + D(x)^{-1}\mu(x)$ denote the circulant [23]. The circulant is the remainder left over when approximating $D(x)^{-1}\mu(x)$ with the negative gradient of the quasipotential. If S_0 satisfies the Hamilton-Jacobi equation then the circulant is perpendicular to the gradient of S_0 under the inner-product weighted by the diffusion tensor:

$$\mu_{circ}(x) \cdot D(x)\nabla S_0(x) = 0.$$

Suppose $D(x)$ is proportional to the identity. Then, since S_0 is a scalar function the vector field Equation (8.41) requires the vector field μ_{circ} moves along isoclines of S_0. For

this reason the circulant is sometimes called a transverse vector field. Notice the strong resemblance to the HHD and Helmholtz potential. Like the HHD we are searching for a scalar potential function whose gradient generates a conservative field that approximates a vector field, and, when the conservative field is removed from the original vector field the remainder field circulates.

The quasipotential has a number of important properties. The most intuitive is proved below:

Lemma 38. *Given an equilibrium distribution $q(x) = \exp S_0(x|\sigma)$ the quasipotential (negated first term in the Laurent expansion of $S_0(x|\sigma)$ in σ^2) is a Lyapunov function for the deterministic eqution $\frac{d}{dt}x = \mu(x)$ [23].*

Proof. Given $\frac{d}{dt}x(t) = \mu(x)$:

$$\begin{aligned}\frac{d}{dt}S_0(x(t)) &= \nabla S_0(x) \cdot \frac{d}{dt}x = \nabla S_0(x) \cdot \mu(x) \\ &= \nabla S_0(x) \cdot D(x)D^{-1}(x)\mu(x) = \nabla S_0(x) \cdot D(x)f(x)\cdot \\ &= \nabla S_0(x) \cdot D(x)(-\nabla S_0(x) + \mu_{circ})\end{aligned}$$

The diffusion tensor $D(x)$ is positive semi-definite for all x so the product $\nabla S_0(x) \cdot D(x)\nabla S_0(x) \geq 0$. By definition of the circulant, the inner product with the circulant $\nabla S_0(x) \cdot D(x)\mu_{circ}$ is zero so $\frac{d}{dt}S(x(t)) \leq 0$.

□

To make the comparison with the Helmholtz potential more direct we define a generalized quasipotential by modifying the vector field appearing in the Hamilton-Jacobi equation

and generalize the inner-product to allow for arbitrary weights:

$$\nabla\phi_q(x) \cdot W(x)(\nabla\phi_q(x) + f(x)) = 0. \tag{8.43}$$

Here $W(x)$ is an arbitrary positive-definite weight matrix, and $f(x)$ are the forces. Note that the forces $f(x) = D^{-1}(x)(\mu(x) - \nabla \cdot D(x))$ only differ from the vector field that appears in the quasipotential by the inclusion of the Itô term $-D^{-1}\nabla \cdot D(x)$. This arises from the zeroeth order term in the Laurent expansion of $S(x)$, and is zeroeth order in σ if $D(x)$ is replaced with $\sigma^2 D(x)$. The first term is $\mathcal{O}(\sigma^{-2})$ so it dominates in the small noise limit. Thus, in the small noise limit the forces converge to the vector field decomposed using the quasipotential. If $W(x)$ is set to $D(x)$ and a small noise limit is taken, then the generalized Hamilton-Jacobi equation converges to the Hamilton-Jacobi equation associated with the the Friedlin-Wentzell quasipotential.

Equation (8.43) allows for easy comparison with the Helmholtz potential. The potentials obey the Poisson equation and Hamilton-Jacobi equations respectively:

$$\begin{aligned} \nabla \cdot W(x) \cdot (\nabla\phi(x) + f(x)) &= 0 \\ \nabla\phi_q(x) \cdot W(x)(\nabla\phi_q(x) + f(x)) &= 0 \end{aligned} \tag{8.44}$$

where the weight matrices may arise from a change of coordinates, or from the diffusion tensor. Note that both equations consist of a requirement on the remainder left over after approximating the forces with the negative gradient of a potential function. The Poisson equation requires that, after weighting, the remainder is incompressible. The Hamilton-Jacobi equation requires that, after weighting, the remainder is orthogonal to the gradient of the potential. While these equations are similar, this last distinction is important. It introduces different geometric requirements on the remainder, and hence what it means to

find a remainder that circulates.

8.4.3 The Effective Potential

Let $q(x)$ be the steady state distribution for the SDE. Then the effective potential is defined:

$$\phi_{\text{eff}}(x) = -\log(q(x)) \tag{8.45}$$

or, in some limiting scenarios, as a function of the limiting parameter times $-\log(q(x))$. For example, if the noise intensity is vanishing then we might define $\phi_{\text{eff}}(x) = -\frac{1}{\sigma^2}\log(q(x))$. These definitions are motivated by the Boltzmann equation. Written this way, the effective potential plays the same role as energy if the system is energetically closed.

We have hinted at the effective potential throughout this discussion, but will now explore it in depth. Like the Helmholtz potential and quasipotential, the effective potential can be expressed as the solution to a PDE involving the approximation of the forces with the gradient of a potential function. In fact, we will show that the PDE governing the effective potential is intimately related to both the Poisson and Hamilton-Jacobi equations. We then use this relation to derive an equivalence theorem, and to show how the potentials are related to the steady state distribution.

To start, consider the case when $D(x) = \sigma^2 I$. Then, as usual, we begin by writing down the stationary Fokker-Planck equation:

$$\frac{d}{dt}\pi(x,t) = -\nabla \cdot (\mu(x)\pi(x,t)) + \sigma^2 \nabla^2 \pi(x,t).$$

Then stationarity requires:

$$-\nabla \cdot (\mu(x)q(x)) + \sigma^2 \nabla^2 q(x) = 0.$$

Then suppose that $q(x)$ takes the form:

$$q(x) = \frac{1}{Z} \exp\left(-S(x)\right).$$

so $S(x)$ is proportional to the effective potential. Then the gradient of the stationary distribution is:

$$\nabla q(x) = \nabla \frac{1}{Z} \exp\left(-S(x)\right) = \left(-\nabla S(x)\right) q(x).$$

and by the divergence product rule:

$$-\nabla \cdot (\mu(x)q(x)) = -\left([\nabla \cdot \mu(x)]q(x) - [\mu(x) \cdot \nabla S(x)]q(x)\right)$$

$$\nabla^2 q(x) = \nabla \cdot \nabla q(x) = \nabla \cdot \left[\left(-\nabla S(x)\right) q(x)\right] = [\nabla S(x) \cdot \nabla S(x)]q(x) - [\nabla^2 S(x)]q(x).$$

Substituting into the stationarity condition and canceling the common factor of $q(x)$:

$$\nabla \cdot \mu(x) - \mu(x) \cdot \nabla S(x) = \sigma^2 \left[\nabla S(x) \cdot \nabla S(x) - \nabla^2 S(x)\right]$$

Divide across by σ^2 and replace $\mu(x)/\sigma^2$ with $f(x)$:

$$\nabla \cdot f(x) - f(x) \cdot \nabla S(x) = \nabla S(x) \cdot \nabla S(x) - \nabla^2 S(x).$$

Finally, rearrange the equation so that the left hand and right hand sides are familiar:

$$\nabla \cdot (\nabla S(x) + f(x)) = \nabla S(x) \cdot (\nabla S(x) + f(x)) \tag{8.46}$$

Equation (8.46) is the necessary requirement for $S(x)$ to be proportional to the effective potential if the noise is isotropic and its instantaneous variance constant. Notice that the left hand side is the divergence of the circulant (the difference between the conservative field and the stochastic field), while the right hand side is the inner product between the conservative field and the circulant. That is, the left hand side is the left hand side of the Poisson equation, and the right hand side is the left hand side of the Hamilton-Jacobi equation. If both are independently zero, then $S(x)$ is automatically proportional to the effective potential. If both sides are independently zero then the circulant is divergence-free and the circulant is orthogonal to the conservative field at every x.

Equation (8.46) can be rewritten:

$$\nabla \cdot (\nabla S(x) + f(x)) - \nabla S(x) \cdot (\nabla S(x) + f(x)) = 0. \tag{8.47}$$

In this form it is clear that the effective potential obeys a PDE that is a mixture of both the Hamilton-Jacobi equation and the Poisson equation which define the quasipotential and Helmholtz potential.

Now suppose that the noise is not isotropic and constant. Then the Fokker-Planck equation takes the more complicated form:

$$\frac{d}{dt}\pi(x,t) = -\nabla \cdot (\mu(x)\pi(x,t)) + \sigma^2 \sum_{ij} \partial_{x_i}\partial_{x_j} D_{ij}(x)\pi(x,t).$$

Which gives the equilibrium condition:

$$\nabla \cdot \left(\left[\mu(x) - \frac{\sigma^2}{2} \sum_j \partial_{x_j} D_{ij}(x) \right] q(x) \right) = \frac{\sigma^2}{2} \nabla \cdot D_{ij}(x) \nabla q(x).$$

Next let:

$$f(x) = \frac{1}{\sigma^2} D^{-1}(x) \left(\mu(x) - \sigma^2 \nabla \cdot D(x) \right)$$

denote the forces. In the small noise limit the forces converge to the vector field $D^{-1}(x)\mu(x)$, which is decomposed into the quasipotential and the circulant when using the Friedlin-Wentzell quasipotential.

Then, applying the divergence and gradient product rules, and canceling the common factor of $q(x)$ from both sides of the equation gives the general stationarity equation for the effective potential:

$$\nabla \cdot D(x) \left(\nabla S(x) + f(x) \right) = \nabla S(x) \cdot D(x) \left(\nabla S(x) + f(x) \right) \tag{8.48}$$

As in the special case when $D(x) = I$, the stationarity condition is a mixture of the Poisson equation and Hamilton-Jacobi equation:

$$\nabla \cdot D(x) \left(\nabla S(x) + f(x) \right) - \nabla S(x) \cdot D(x) \left(\nabla S(x) + f(x) \right) = 0. \tag{8.49}$$

Equation (8.48) can be used to work out the large and small noise limits of the effective potential. Here we focus on the constant isotropic noise case for simplicity. We have already shown that the quasipotential is a small noise limit of the effective potential, if we

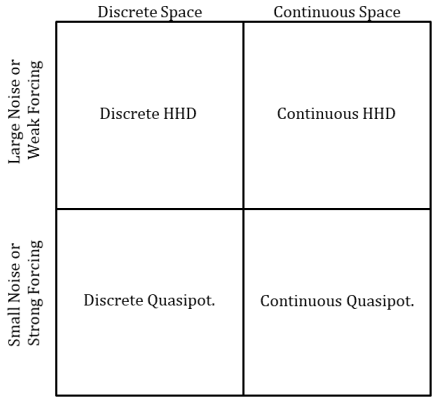

Figure 8.4: Classification of which potential to use, given state space and noise/forcing limit.

assume the effective potential has the form:

$$S(x, \sigma^2) = \frac{1}{\sigma^2} \sum_{m=0}^{\infty} S_m(x) \sigma^{2m}.$$

Then:

$$\lim_{\sigma^2 \to 0} \sigma^2 \left[\nabla \cdot (S(x) + f(x)) + \nabla S(x) \cdot (S(x) + f(x)) \right] = \lim_{\sigma^2 \to 0} \mathcal{O}(\sigma^0) + \mathcal{O}(\sigma^{-2}) = 0$$

is dominated by the second half of the bracketed expression since it depends on an additional factor of σ^{-2}. Thus the limit leaves the Hamilton-Jacobi equation for the quasipotential.

The same technique can be used to isolate the Poisson term if we take a large noise limit. In the large noise limit the stationary distribution becomes increasingly smooth. When the noise is large the forces are small, so a large noise limit is analogous to the weak forcing limits considered in Section 7.3. When the noise is large, diffusion dominates drift. If the

noise is isotropic and constant, then the steady state approaches a uniform distribution. It follows that the effective potential should approach an arbitrary constant, which we pick to be zero. The constant is arbitrary since the distribution is normalized by Z. Then, it is natural to expand the effective potential for large σ^2 as:

$$S(x,\sigma) = \frac{1}{\sigma^2}\sum_{m=0}^{\infty} S_m(x)\sigma^{-2m}.$$

Then:

$$\lim_{\sigma^2\to\infty} \sigma^2\left[\nabla\cdot(S(x)+f(x)) + \nabla S(x)\cdot(S(x)+f(x))\right] = \lim_{\sigma^2\to\infty} \mathcal{O}(\sigma^0) + \mathcal{O}(\sigma^{-2}) = 0$$

so the Poisson term dominates. Therefore:

$$\phi(x) \propto \lim_{\sigma^2\to\infty} \sigma^2\phi_{\text{eff}}(x). \tag{8.50}$$

It follows that in the large noise limit, if the diffusion tensor is constant and isotropic, then the effective potential converges to the Helmholtz potential, and in the small noise limit converges to the quasipotential. Note that this limiting behavior mimics the same results observed for discrete space processes, and is natural given the different path integral interpretations of the Helmholtz potential and quasipotential (see Figure 8.4).

8.5 Comparison

We are now equipped to compare the three potentials. The three potentials (Helmholtz, quasi, effective) all decompose the forces. Each attempts to approximate the forces with the negative gradient of a potential function, and enforces a requirement on the error

in that approximation. The remainder left over after approximating the forces with the gradient of a potential is a circulant. The Helmholtz potential requires that the circulant is incompressible. The quasipotential requires that the circulant is transverse (orthogonal to the gradient of the quasipotential).[6] The effective potential mixes the two requirements.

The three potentials are defined by the Poisson, Hamilton-Jacobi, and stationarity equations respectively:

$$\begin{aligned}
&\textbf{Poisson: } \nabla \cdot D(x)(\nabla\phi(x) + f(x)) = 0 \\
&\textbf{Hamilton-Jacobi: } \nabla\phi_q(x) \cdot D(x)(\nabla\phi_q(x) + f(x)) = 0 \\
&\textbf{Stationarity: } (\nabla - \nabla\phi_{\text{eff}}(x)) \cdot D(x)(\nabla\phi_{\text{eff}}(x) + f(x)) = 0
\end{aligned} \tag{8.51}$$

Equation (8.51) lays the groundwork for analyzing cases when the potentials are all equivalent. These are explored in the next section.

8.5.1 Conditions for Equivalence

Theorem 39 (Potential Equivalence)**.** *Given an SDE $X(t)$ in $\mathbb{R}^n$ defined by Equation (8.3) with a diffusion tensor $D(x)$ that is invertible for all x that can be reached by $X(t)$ with nonzero probability, then either all three potentials (Helmholtz, quasi, effective) can be chosen so that they are equivalent, or none of the three potentials are equivalent.*

Proof. Either none of the three potentials are equivalent, exactly two are equivalent, or all three are equivalent. We will prove that it is impossible for exactly two to be equivalent. Suppose two of the potentials are equivalent. Then there is a function $S(x)$ that satisfies

[6] Notice that it is possible to be transverse and compressible. For example, if the isoclines of the potential are circular and centered at the origin, and the circulant follows the isoclines, but follows them clockwise for negative x and counterclockwise for positive x, then the circulant is transverse, but is compressible since the divergence is nonzero along $x = 0$.

two of the equations in Equation (8.51). But if $S(x)$ satisfies two of the equations it must satisfy the third equation, since any of the three equations can be expressed as a linear combination of the other two equations. Thus, if two of the potentials are equivalent there exists a potential that satisfies all three equations.

□

Theorem 39 is useful since it is often easier to solve and check the Poisson and Hamilton-Jacobi equations than the stationarity equations. If a solution to the Poisson equation can be identified which satisfies the Hamilton-Jacobi equation then it is necessarily also the effective potential. Alternatively, in cases when the effective potential is known, the equivalence theorem can be used to check whether it is also a Helmholtz and quasipotential without checking both the Poisson and Hamilton-Jacobi equations.

Theorem 39 can be used to derive other equivalence requirements. For example, suppose that $X \in \mathbb{R}^2$ and $D(x) = I$. The HHD requires that the circulant can be written as the curl of a scalar valued function $\theta(x)$. In two dimensions the curl is simply a ninety-degree rotation R of the gradient. Therefore the circulant must be expressible as $R\nabla\theta(x)$. The Hamilton-Jacobi equation is now:

$$(\nabla S(x))^T R(\nabla\theta(x)) = 0.$$

For any given x this takes the form $v^T Rw = 0$ for some pair of vectors v, w. Since R rotates by ninety-degrees, v is orthogonal to Rw if and only if v is parallel to w. Therefore the Hamilton-Jacobi equation requires that the vector potential is constant along the isoclines of $S(x)$. Therefore the vector potential must be some scalar valued function of the scalar potential:

$$\theta(x) = \Lambda(S(x)). \tag{8.52}$$

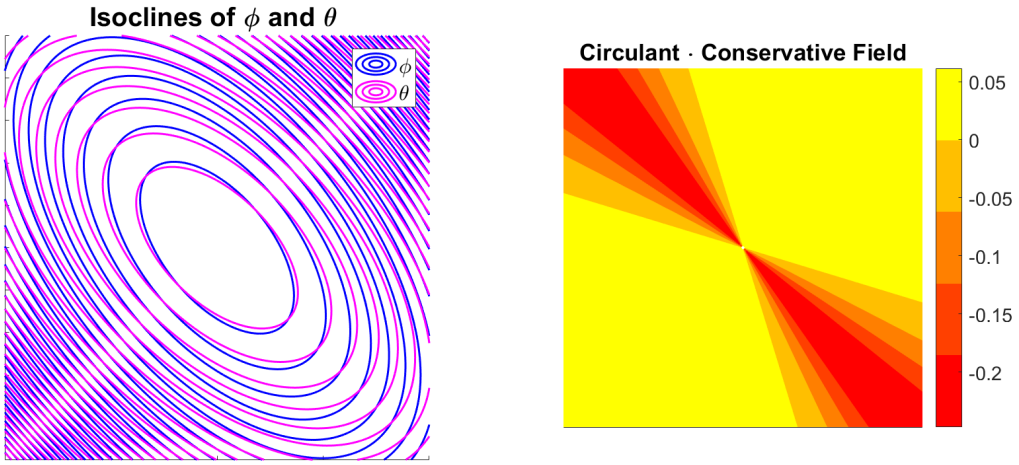

Figure 8.5: Isoclines of θ and ϕ in 2D are shown on the left. Since these isoclines are not parallel the inner product between the circulant and the conservative field is not zero everywhere as illustrated in the right hand panel.

Then, by the chain rule:

$$\nabla\theta(x) = [\partial_s \Lambda(S(x))]\nabla S(x)$$

so the Hamilton-Jacobi equation is satisfied automatically. Thus, in two dimensions, and for constant isotropic noise, the potentials are only equivalent if the vector potential can be expressed as a scalar function of the scalar potential (Helmholtz). Therefore, the Helmholtz potential is only the effective potential for a two-dimensional process with constant isotropic noise if the vector potential is a scalar function of the scalar potential. An example is illustrated in Figure 8.5 where the isoclines of the two potential functions are not parallel, so the Helmholtz potential is not equivalent to the quasipotential or effective potential.

More generally:

Corollary 39.1 (Equivalence in Detailed Balance)**.** *If $X(t)$ is governed by an SDE that obeys detailed balance (forces $f(x)$ are conservative) then all three potentials are equiva-*

lent.

Proof. If the SDE obeys detailed balance then the forces $f(x)$ are conservative so $f(x) = -\nabla S(x)$ for some $S(x)$. Then if all three potentials are set equal to $S(x)$ the Poisson, Hamilton-Jacobi, and stationarity equations are all satisfied simultaneously. □

Outside of detailed balance the three potentials are usually not equivalent. However, in some important special cases all three are the same. In the next section we will show that if the SDE is an Ornstein-Uhlenbeck (OU) process then it is always possible to find a potential function which is simultaneously a Helmholtz, quasi, and effective potential.

8.5.2 Equivalence for OU Processes

An Ornstein-Uhlenbeck (OU) process is an SDE of the form:

$$dX = -AX(t)dt + BdW. \tag{8.53}$$

OU processes are widely used to approximate the behavior of SDEs near a stable attractor of the deterministic process, or fluctuations away from deterministic trajectories [281]. OU processes are also widely used to model stochastic oscillators. Throughout this section we will assume that A is positive definite and $D = \frac{1}{2}BB^T$ is invertible. These assumptions guarantee that there is a Gaussian steady state with mean zero and finite covariance.

To start, consider an OU process in $\mathbb{R}^2$ centered at $x_0 = 0$ with isotropic noise. Then let $\mu(x) = -Ax$ and $D(x) = I$ so:

$$dX = -AXdt + \sigma dW. \tag{8.54}$$

Since the domain is unbounded and $\mu(x)$ is not L_1 integrable the HHD of the stochastic field is *not unique* [8]. It follows that there is a space of potentials ϕ, θ such that:

$$-\nabla\phi(x) + \nabla \times \theta(x) = \frac{1}{\sigma^2}\mu(x) = -\frac{1}{\sigma^2}Ax. \tag{8.55}$$

Our goal is to find a particular pair of potentials ϕ, θ such that the scalar potential ϕ is equivalent to the effective potential for the system, and converges to the quasipotential in the small noise limit. This requires that the potentials also satisfy the Hamilton-Jacobi equation:

$$\nabla\phi \cdot \left(\nabla\phi + \frac{1}{\sigma^2}\mu(x)\right) = (\nabla\phi) \cdot (\nabla \times \theta(x)) = 0. \tag{8.56}$$

That is, the conservative field $\nabla\phi$ must be orthogonal to the rotational field $\nabla \times \theta$ at every x.

The main advantage of working in 2D is that the curl operator is just a rotation of the gradient operator. This simple relation allows us to solve explicitly for ϕ and θ in terms of the entries of A.

The gradient is:

$$\nabla = \begin{bmatrix} \partial_{x_1} \\ \partial_{x_2} \end{bmatrix}$$

and the curl is:

$$\nabla\times = \begin{bmatrix} \partial_{x_2} \\ -\partial_{x_1} \end{bmatrix}.$$

Therefore, if we define the rotation matrix R that rotates each vector by ninety degrees:

$$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then:

$$\nabla\times = R\nabla.$$

The right hand side of the HHD is linear, therefore we expect both potentials to be quadratic, and centered at the origin:

$$\begin{aligned}\phi(x) &= \frac{1}{2}xBx^T \\ \theta(x) &= \frac{1}{2}xCx^T\end{aligned} \tag{8.57}$$

for some pair of matrices B and C. Any quadratic form is symmetric, for example:

$$x^TBx = b_{11}x_1^2 + (b_{12}+b_{21})x_1x_2 + b_{22}x_2^2 = b_{11}x_1^2 + (b_{21}+b_{12})x_1x_2 + b_{22}x_2^2.$$

Therefore, without loss of generality we assume that both B and C are symmetric.

In general, for any symmetric matrix:

$$\nabla\frac{1}{2}xBx^T = Bx.$$

So:

$$\nabla\times\frac{1}{2}xCx^T = RCx.$$

Therefore:

$$-\nabla\phi(x) + \nabla\times\theta(x) = (-B+RC)x.$$

Plugging into the HHD we get the linear equation:

$$(-B+RC)x = Ax. \tag{8.58}$$

Since B and C are both symmetric the left hand side of this equation has six degrees of freedom. The matrix A has four degrees of freedom, so there is a two dimensional space of possible B and C. In order to solve for some constraints on B and C take the divergence and curl of the HHD to get a pair of Poisson's equation. Since the divergence is orthogonal to the curl, and the curl to the gradient:

$$
\begin{aligned}
-\nabla^2\phi(x) &= -\nabla\cdot Bx = -\text{trace(B)} = \nabla\cdot Ax = \text{trace(A)} \\
-\nabla^2\theta(x) &= \nabla\times RCx = \text{trace(C)} = \nabla\times Ax = \text{trace(R A)}.
\end{aligned}
$$

That is:

$$
\begin{aligned}
\text{trace(B)} &= b_{11}+b_{22} = -\text{trace(A)} - (a_{11}+a_{22}) \\
\text{trace(C)} &= c_{11}+c_{22} = \text{trace(R A)} = a_{12}-a_{21}.
\end{aligned}
\tag{8.59}
$$

Clearly this is not enough to specify B and C. If we treat the trace of the matrix appearing in a quadratic form as a general measure of the size of the associated quadratic potential then the first equation requires that the scalar potential is about the same size as the diagonal of A, while the second equation requires that vector potential is about the same size as the asymmetric part of A. In fact, if A is symmetric then the system is conservative and we can set $B = A$ and $C = 0$. More generally, if A is not symmetric we could let:

$$
\begin{aligned}
B &= -\frac{1}{2}\left(A+A^T\right) = \begin{bmatrix} a_{11} & \frac{1}{2}(a_{12}+a_{21}) \\ \frac{1}{2}(a_{12}+a_{21}) & a_{22} \end{bmatrix} \\
C &= \frac{1}{2}R^T\left(A-A^T\right) = \begin{bmatrix} \frac{1}{2}(a_{12}-a_{21}) & 0 \\ 0 & \frac{1}{2}(a_{12}-a_{21}) \end{bmatrix}.
\end{aligned}
\tag{8.60}
$$

Then $-\nabla\phi(x) = \frac{1}{2}\left(A+A^T\right)x$ and, since $RR^T = I$, $\nabla\times\theta(x) = R\frac{1}{2}R^T\left(A-A^T\right) = \frac{1}{2}(A-A^T)x$. Then, $-\nabla\phi+\nabla\times\theta = \frac{1}{2}\left(A+A^T\right)x + \frac{1}{2}(A-A^T)x = Ax$.

That said, this decomposition does not satisfy the Hamilton-Jacobi equation since:

$$\left(\frac{1}{2}\left(A+A^T\right)x\right)\cdot\left(\frac{1}{2}\left(A-A^T\right)x\right)=\frac{1}{4}\left(x^TAAx+x^TA^TAx-x^TAA^Tx-x^TA^TA^Tx\right)$$

and:

$$\frac{1}{4}\left(x^TAAx+x^TA^TAx-x^TAA^Tx-x^TA^TA^Tx\right)=\frac{1}{4}x^T(A^TA-AA^T)x\neq 0$$

unless A is a normal matrix (is unitarily diagonalizable).

In order to find a decomposition that does satisfy the Hamilton-Jacobi equation rewrite the Hamilton-Jacobi in terms of R:

$$(\nabla\phi(x))^TR(\nabla\theta(x))=0.$$

In general the quadratic form $v^TRw=0$ if and only if v is parallel to w. Therefore the Hamilton-Jacobi equation requires:

$$\nabla\theta(x)=s(x)\nabla\phi(x) \tag{8.61}$$

for some scalar function $s(x)$. For OU processes it is sufficient to assume $s(x)$ is a constant s. Then, plugging into the left hand side of the HHD:

$$-\nabla\phi(x)+\nabla\times\theta(x)=-\nabla\phi(x)+R\nabla\theta(x)=-\nabla\phi(x)+sR\nabla\phi(x)=(-I+sR)\nabla\phi$$

which gives:

$$(-I+sR)\nabla\phi(x)=\frac{1}{\sigma^2}Ax \tag{8.62}$$

where both s and $\phi(x)$ are unknown. Substitute in the quadratic form for $\phi(x)$:

$$(-I + sR)Bx = \frac{1}{\sigma^2}Ax \tag{8.63}$$

or:

$$(-I + sR)B = \frac{1}{\sigma^2}A. \tag{8.64}$$

For simplicity assume $\frac{1}{\sigma^2} = 1$. We can generalize our results to arbitrary σ by multiplying by $\frac{1}{\sigma^2}$.

Now the goal is to solve for the four unknowns $b_{11}, b_{12} = b_{21}, b_{22}$ and s given the four knowns $a_{11}, a_{12}, a_{21}, a_{22}$. As long as s is unknown this is not a linear problem. To solve, write the four equations explicitly (noting $b_{12} = b_{21}$):

$$\begin{aligned} -b_{11} + sb_{12} &= a_{11} \\ -b_{12} + sb_{22} &= a_{12} \\ -sb_{11} - b_{12} &= a_{21} \\ -sb_{12} - b_{22} &= a_{22}. \end{aligned} \tag{8.65}$$

Notice that if we add the first and fourth equation we get:

$$-b_{11} + sb_{12} - sb_{12} - b_{22} = -\text{trace}(B) = a_{11} + a_{22} = \text{trace}(A)$$

which matches the requirement we derived from Poisson's equations:

$$\text{trace}(B) = -\text{trace}(A) \tag{8.66}$$

Next note that if we subtract the third equation from the second equation we get:

$$-b_{12} + sb_{22} + sb_{12} + b_{12} = s(b_{11} + b_{22}) = s\text{trace}(B) = a_{12} - a_{21}.$$

Plugging in $\text{trace}(B) = -\text{trace}(A)$ and simplifying we find:

$$s = -\frac{a_{12} - a_{21}}{\text{trace}(A)}. \tag{8.67}$$

That is, the vector potential is proportional to the scalar potential with a factor s that is the ratio between the asymmetric part of A and its trace. Notice that this proportionality corresponds to the requirement on the vector potential we derived from Laplace's equation. Also notice that if A is symmetric then s is zero.

Now that s is known the four equations are a linear system in three unknowns. In order to solve this linear system it is convenient to let $b_{11} = -\frac{1}{2}\text{trace}(A) + u$ and $b_{22} = -\frac{1}{2}\text{trace}(A) - u$. Then:

$$\begin{aligned}
&\frac{1}{2}(a_{11} + a_{22}) - u + sb_{12} = a_{11} \\
&-b_{12} - \frac{a_{12} - a_{21}}{\text{trace}(A)}\left(-\frac{1}{2}\text{trace}(A) - u\right) = a_{12} \\
&\frac{a_{12} - a_{21}}{\text{trace}(A)}\left(-\frac{1}{2}\text{trace}(A) + u\right) - b_{12} = a_{21} \\
&-sb_{12} + \frac{1}{2}(a_{11} + a_{22}) + u = a_{22}.
\end{aligned}$$

To simplify first cancel the common factor of the trace of A where possible, and subtract

all terms in the first and fourth equations involving a_{11} and a_{22} to the right hand side:

$$\begin{aligned}
-u + sb_{12} &= \frac{1}{2}(a_{11} - a_{22}) \\
-b_{12} + \frac{1}{2}(a_{12} - a_{21}) - su &= a_{12} \\
-\frac{1}{2}(a_{12} - a_{21}) - su - b_{12} &= a_{21} \\
-sb_{12} + u &= -\frac{1}{2}(a_{11} - a_{22}).
\end{aligned}$$

Notice that the first and fourth equations are identical. This should not be a surprise since we have already solved for two unknowns. If we subtract all terms in equations two and three involving a_{12} and a_{21} to the left hand side then the two equations are identical, leaving the linear system:

$$\begin{aligned}
-u + sb_{12} &= \frac{1}{2}(a_{11} - a_{22}) \\
-b_{12} - su &= \frac{1}{2}(a_{12} + a_{21})
\end{aligned}$$

Notice that the two terms left on the right hand side correspond to the difference along the diagonal of A and the sum off the off-diagonal elements. This should not be a surprise since we have already ensured the decomposition would match the sum of the diagonal elements of A and the difference of the off-diagonal elements.

If A is symmetric then s is zero, therefore:

$$\begin{aligned}
u &= -\frac{1}{2}(a_{11} - a_{22}) \\
b_{12} &= -\frac{1}{2}(a_{12} + a_{21})
\end{aligned}$$

which automatically gives $B = -A$.

If A is not symmetric then s is not zero so we need to solve the linear system for u and

b_{12}. The linear system has solution:

$$\begin{aligned} u &= -\frac{1}{2}\frac{1}{(s^2+1)}\left[(a_{11}-a_{22})+s(a_{12}+a_{21})\right] \\ b_{12} &= \frac{1}{2}\frac{1}{(s^2+1)}\left[s(a_{11}-a_{22})-(a_{12}+a_{21})\right] \end{aligned} \tag{8.68}$$

Now plugging back in for b_{11} and b_{22} we find:

$$\begin{aligned} b_{11} &= -\frac{1}{2}(a_{11}+a_{22}) - \frac{1}{2}\frac{1}{(s^2+1)}\left[(a_{11}-a_{22})+s(a_{12}+a_{21})\right] \\ b_{12} = b_{21} &= \frac{1}{2(s^2+1)}\left[s(a_{11}-a_{22})-(a_{12}+a_{21})\right] \\ b_{22} &= -\frac{1}{2}(a_{11}+a_{22}) + \frac{1}{2}\frac{1}{(s^2+1)}\left[(a_{11}-a_{22})+s(a_{12}+a_{21})\right] \end{aligned} \tag{8.69}$$

As usual, if $s = 0$ then we get $B = -A$.

If s is small then:

$$B = -\frac{1}{2}[A + A^T] - s\frac{1}{2}\begin{bmatrix} (a_{12}+a_{21}) & (a_{22}-a_{11}) \\ (a_{22}-a_{11}) & -(a_{12}+a_{21}) \end{bmatrix} + \mathcal{O}(s^2). \tag{8.70}$$

Note that this differs from the most natural decomposition $B = -\frac{1}{2}[A + A^T]$ by the factor $-\frac{1}{2}s[RA - AR]$.

In summary, given an OU process in $\mathbb{R}^2$ with deterministic skeleton $\mu(x) = -Ax$ the scalar potential and vector potentials take the forms $\phi(x) = \frac{1}{2}x^T Bx$ and $\theta(x) = s\phi(x)$ where $s = -\frac{a_{12}-a_{21}}{\text{trace}(A)}$ and B is symmetric with b_{11}, b_{12}, and b_{22} as given by Equation (8.69). Then $\phi(x), \theta(x)$ are an HHD of the forces and satisfy the Hamilton-Jacobi equation. It follows that the rotational field is orthogonal to the conservative field everywhere and $\phi(x)$ is equivalent to both the effective potential and the quasipotential. Therefore, there exists at least one generic non-detailed balance case where the potentials can be chosen

so that the Helmholtz potential is equivalent to the effective potential and the quasipotential.

We have also shown that for a two dimensional OU process there exists a natural measure of how much rotation is present in a system relative to its conservative part, s, where s is given by the ratio of the asymmetric part of A to its trace. It is worth noting that $|s|> 0$ whenever A is not symmetric, so even if A has all real eigenvalues the field Ax is not necessarily conservative, and the vector potential may not be zero.[7]

These results can be generalized to higher dimension and to anisotropic noise. Before considering the general case we can easily extend our first result from the previous section, that is, if the matrix A is normal and noise is isotropic then $\phi(x) = \frac{1}{4}x^T(A + A^T)x$ is equivalent to all three potentials.

As in the previous section assume that $X(t)$ is a stochastic process which takes values on $\mathbb{R}^n$ with deterministic skeleton $\mu(x) = -Ax$ and diffusion $D(x) = I$. Also assume that A is normal:

$$A^T A = AA^T.$$

Then:

Lemma 40. *Given any OU process in $\mathbb{R}^n$ with drift $\mu(x) = -Ax$ and diffusion tensor $D = \sigma^2 I$ then, if A is normal, there exists a potential function:*

$$S(x) = \frac{1}{4\sigma^2}x^T(A + A^T)x \tag{8.71}$$

that is equal to the Helmholtz potential, quasipotential, and effective potential up to the addition of a constant.

[7] The ratio s is also the area production defined by [281, 282].

Proof. The forces are:

$$f(x) = -\frac{1}{\sigma^2}Ax = -\frac{1}{2\sigma^2}\left(A + A^T\right)x - \frac{1}{2\sigma^2}\left(A - A^T\right)x.$$

Let:

$$S(x) = \frac{1}{4\sigma^2}x^T(A + A^T)x$$

Then:

$$-\nabla S(x) = -\frac{1}{2\sigma^2}(A + A^T)x.$$

Therefore:

$$\nabla S(x) + f = -\frac{1}{2\sigma^2}\left(A - A^T\right)x.$$

In general the $\nabla \cdot Bx$ is the trace of B. The matrix $A - A^T$ is all zero along its diagonal so:

$$\nabla \cdot (\nabla S(x) + f(x)) = -\nabla \cdot \frac{1}{2\sigma^2}\left(A - A^T\right)x = -\frac{1}{2\sigma^2}\text{trace}(A - A^T) = 0.$$

Therefore $S(x)$ satisfies the Poisson equation, so is a Helmholtz potential.

Next, check the Hamilton-Jacobi equation:

$$\nabla S(x) \cdot (\nabla S(x) + f(x)) = \frac{1}{4\sigma^4}x^T\left(A + A^T\right)\left(A - A^T\right)x.$$

Expanding:

$$x^T\left(A + A^T\right)\left(A - A^T\right)x = x^TAAx - x^TAA^Tx + x^TA^TAx - x^TA^TA^Tx.$$

But, $x^T AAx = x^T A^T A^T x$ so:

$$\nabla S(x) \cdot (\nabla S(x) + f(x)) = \frac{1}{4\sigma^4} x^T (A^T A - AA^T) x.$$

For general A the commutator $A^T A - AA^T$ is not zero. However, if A is normal then $A^T A = AA^T$ by definition so:

$$\nabla S(x) \cdot (\nabla S(x) + f(x)) = 0.$$

It follows that $S(x)$ satisfies the Hamilton-Jacobi equation so is proportional to the quasipotential. Then, by the equivalence relation between the potentials, $S(x)$ must also be proportional to the effective potential. □

More generally, the stationary distribution for any OU process in $\mathbb{R}^d$ with diffusion tensor D is a multivariate normal distribution with mean zero:

$$q(x) \propto \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \tag{8.72}$$

with covariance Σ. It follows that the effective potential always takes the form:

$$\phi_{\text{eff}}(x) = \frac{1}{2} x^T \Sigma^{-1} x. \tag{8.73}$$

The covariance matrix is symmetric, and is the solution to the Lyapunov equation [75]:

$$A\Sigma + \Sigma A^T = 2D. \tag{8.74}$$

Using the Fredholm alternative it can be shown that the Lyapunov equation has a unique

solution in the space of symmetric matrices [71].

By Theorem 39, if we can show that the effective potential satisfies the weighted Poisson equation $\nabla \cdot D(x) \left(\nabla \phi_{\text{eff}}(x) + f(x)\right) = 0$ then ϕ_{eff} satisfies the Hamilton-Jacobi equation automatically. Therefore, to establish general equivalence we need to show that:

$$\nabla \cdot D \left(\Sigma^{-1} x - D^{-1} A x\right) = 0$$

The divergence of a matrix times x is the trace of the matrix, so the Poisson equation reduces to:

$$\text{trace}[D\Sigma^{-1}] = \text{trace}[A].$$

By exploiting the Lyapunov equation and the invertibility of Σ we can prove that this is true, and thus, for any OU process there exists a potential which is simultaneously a Helmholtz, quasi, and effective potential.

Theorem 41. *Given an OU process in $\mathbb{R}^n$ with drift $\mu(x) = -Ax$ where A is positive definite, and with a positive definite diffusion tensor D, there exists a potential function $S(x)$ that is proportional to the Helmholtz, quasi, and effective, and potentials. The potential is:*

$$S(x) = \frac{1}{2} x^T \Sigma^{-1} x \tag{8.75}$$

where Σ is the covariance matrix for the stationary distribution of the OU process satisfying:

$$A\Sigma + \Sigma A^T = 2D \tag{8.76}$$

Proof. Given $\mu(x) = -Ax$ the forces are given by $f(x) = -D^{-1}Ax$. Theorem 39 states that if there is a potential $S(x)$ that is simultaneously the effective potential and satisfies the Poisson equation then it also satisfies the Hamilton-Jacobi equation and all the potentials

are equivalent. Here we take advantage of the fact that the effective potential is known for OU processes, so we can check whether or not it satisfies the Poisson equation directly.

Given an OU process with positive definite A the stationary distribution is Gaussian with form:

$$q(x) \propto \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right) \tag{8.77}$$

where:

$$A\Sigma + \Sigma A^T = 2D.$$

Therefore let:

$$S(x) = -\log\left(q(x)\right) = \frac{1}{2}x^T\Sigma^{-1}x. \tag{8.78}$$

The covariance matrix Σ is symmetric so Σ^{-1} is also symmetric. Therefore:

$$\nabla S(x) = \Sigma^{-1}x. \tag{8.79}$$

It follows that the Poisson equation is satisfied if:

$$\nabla \cdot D(\Sigma^{-1}x - \frac{2}{\sigma^2}Ax) = \text{trace}\left(D\Sigma^{-1}\right) - \text{trace}\left(A\right) = 0$$

or:

$$\text{trace}\left(D\Sigma^{-1}\right) = \text{trace}\left(A\right).$$

Multiply the Lyapunov equation $A\Sigma + \Sigma A^T = \sigma^2 I$ from the right by Σ^{-1}. Then:

$$A + \Sigma A^T\Sigma^{-1} = 2D\Sigma^{-1}.$$

Take the trace on both sides:

$$\text{trace}(2D\Sigma^{-1}) = \text{trace}(A) + \text{trace}(\Sigma A^T \Sigma^{-1}).$$

But $\Sigma A^T \Sigma^{-1}$ is a similarity transform of A^T so $\text{trace}(\Sigma A^T \Sigma^{-1}) = \text{trace}(A^T)$. The transpose of A has the same diagonal as A so $\text{trace}(A^T) = \text{trace}(A)$. Therefore:

$$\text{trace}\left(2D\Sigma^{-1}\right) = 2\text{trace}\left(A\right). \tag{8.80}$$

or $\text{trace}\left(D\Sigma^{-1}\right) = \text{trace}\left(A\right)$.

Therefore the effective potential $S(x)$ satisfies the Poisson equation hence it must also satisfy the Hamilton-Jacobi equation, so all three potentials are equivalent.

□

Therefore, for any OU process, the effective potential satisfies both the Poisson and Hamilton-Jacobi equation. The general equivalence between the three potentials for OU processes in $\mathbb{R}^n$ leads to general equivalence near stable equilibria of the deterministic skeleton.

Suppose the point x_* is a stable equilibrium of $\mu(x)$. Then $\mu(x_*) = 0$ and we could write:

$$\mu(x_* + y) = -A(x_*)y + \mathcal{O}(y^2) \tag{8.81}$$

where A is the Jacobian matrix of $\mu(x)$ evaluated at x_*. If the diffusion tensor is constant then the linearized SDE is an OU process, and the corresponding potentials are all equivalent. Therefore, given an arbitrary $\mu(x)$, if $\mu(x)$ has stable equilibria it is possible to pick a Helmholtz potential ϕ which converges to the effective potential and the quasipotential at at least one stable equilibrium of the deterministic process.

It also follows that substantive differences in the potentials (differences that cannot be resolved by changing which component of the harmonic field is associated with ϕ) are the result of nonlinear features of $f(x)$. Since we generally want the potentials to agree in the vicinity of stable equilibria, this means that the potentials will primarily differ in their treatment of saddles, and the curvature of each well away from its basin.

8.6 Summary

In this chapter we have illustrated how the discrete HHD developed in Chapter 6 can be extended to the continuum. We then showed that, as for discrete-space processes, the Helmholtz potential is associated with steady state dynamics when the process is dominated by diffusion. We show that the effective potential is governed by a PDE which combines the PDE defining the Helmholtz potential, and the PDE defining the quasipotential, and that one term dominates when noise is small, and the other when noise is large. A general equivalence theorem is presented in Section 8.5, and it is shown that the potentials may be equivalent even if the underlying process does not obey detailed balance.

Part V

Discussion

Chapter 9

Discussion and Future Work

In this dissertation we have shown that the discrete HHD defined by Lim and Jiang [15, 16] is a powerful tool for analyzing edge flows on networks that arise in applications. We illustrate that, if the appropriate edge flow is chosen, then the HHD can be used to describe structure, and to analyze dynamics. Chapters 4 and 5 showed that the HHD can be used to describe the structure of competition in tournaments. Chapters 6 and 7 demonstrate that, when applied to a Markov chain, the HHD can be used to build thermodynamic analogies, is intimately related to nonequilibrium steady states, steady state fluxes, and observable production in the weak rotation limit, and is a complementary potential decomposition to the quasi-potential, which can be generalized to networks.

The results described here point to a variety of interesting avenues for future research.

The methods presented in Chapter 3 could be extended by considering other graph operations (unions, intersections, contractions, strong and tensor products). Cycle basis optimization methods could be implemented to minimize the length of the cycles in the cycle basis, improve the conditioning of the curl, and reduce the reuse of edges.

The trait-performance theorem presented in Chapter 4 could be extended by weakening the statistical assumptions. In particular the assumption that the traits are drawn independently from the network structure could be generalized. Possible modifications were outlined at the end of Chapter 4. An important alternative approach would be to consider other generic covariance structures for the edge flow.

The estimation tools developed in Chapter 5 could be applied to other examples. We did not report all examples tested in this dissertation, and have a library of exciting examples to consider in the future. These include a variety of political examples collated at preflib.org. The examples include: Irish election data from with ranked votes from 44,000 to 64,000 voters on 12-14 candidates [283], 86 elections held by non-profit organizations, trade unions, and professional organizations [175], the 2007 Glasgow city council elections for 21 wards with 5,000 to 13,000 voters and 8 to 13 candidates, the 2006 and 2009 mayoral elections in Burlington Vermont, local elections in Aspen, Berkely, Minneapolis, Oakland, San Francisco, San Leandro, and Takoma Park, the 2002 French Presidential election [284], and elections to the American Psychological Association between 1998 and 2009 (5 candidates and 13,000-20,000 voters [171]. These data sets are both exciting and diverse. Some include actual ranked votes submitted, so the methods could be applied without inferring voter preferences from thermometer polling. Almost all include large sample sizes on the order of 1,000 to 10,000 voters. An interesting extension to both Chapter 4 and Chapter 5 would be to develop a trait-performance model for voter opinion. This could be informed by polling data from the American National Election Study (ANES), which asks respondents to answer a variety of policy and demographic questions in addition to rating candidates.

Part IV developed the theoretical tools necessary to analyze dynamics of Markov processes with the HHD. There is tremendous promise for future work here, since other

dynamical properties could be considered (i.e. mixing times, passage times, and quasi-stability) that are not addressed, and since Part III does not address a specific model system. In the future we plan to apply the HHD to birth-death models of competition between species to compare the results of using the discrete HHD to similar potential analyses using the quasi-potential framework. We have also made preliminary investigations into generalizations of the evolutionary models proposed by [285]. We plan to apply the weak rotation expansion to understand evolutionary dynamics when selection is close to conservative. This analysis would extend existing work relating statistical physics and evolution [286, 287, 288].

In sum, we hope that the work presented here will provide a unified analytic framework for understanding similar problems that arise in diverse fields.

Part VI

Appendices

Appendix A

Estimation Details

A.1 Model

Consider a tournament consisting of m competitors connected by E edges. Assume that the tournament is connected, and that all competition events are purely pairwise. Index the edges from 1 to E. For each edge assign an arbitrary direction. Let $i(k), j(k)$ denote the start and endpoint of edge k. Let p_k denote the probability competitor $i(k)$ beats competitor $j(k)$. Assume that the outcomes of the competition events are independent, and that p_k do not change in time.

Let n_k be the number of competition events observed on edge k. Let W_k be the number of wins observed.

A.2 Likelihood and Prior

The number of wins W_k is binomially distributed:

$$W_k \sim \text{binomial}(n_k, p_k) \tag{A.1}$$

so:

$$\Pr\{W_k = w\} = \left(\begin{array}{c} n_k \\ w \end{array} \right) p_k^w (1 - p_k)^{n_k - w}. \tag{A.2}$$

Then $\mathbb{E}[W_k] = p_k n_k$ and $\mathbb{V}[W_k] = p_k(1 - p_k)n_k$. Order h central moments are $\mathcal{O}(n_k^{\lceil h/2 \rceil})$.

Assume that the win probabilities p are themselves realizations of a random variable. Therefore denote the win probabilities P. Then the likelihood $P_k = p$ given $W_k = w_k$ is:

$$\Pr\{P_k = p | W_k = w_k\} = \left(\begin{array}{c} n_k \\ w_k \end{array} \right) p_k^w (1 - p)^{n_k - w_k}. \tag{A.3}$$

Therefore the likelihood $P_k = p$ follows a Beta distribution:

$$X \sim \text{Beta}(\alpha, \beta) \text{ then } \pi_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \tag{A.4}$$

where $\pi_X(x)$ is the probability density of X, and X is supported on $[0, 1]$. The normalizing factor $B(\alpha, \beta)$ is:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \tag{A.5}$$

$\Gamma(x)$ is the continuous extension of the factorial: $\Gamma(n + 1) = n!$ for any integer n. Thus $1/B(\alpha, \beta)$ is, in effect, a continuous extension of the binomial coefficient. The beta

distribution has moments:

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\alpha}{\alpha+\beta} \\
\mathbb{V}[X] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
\mathbb{E}[\ln(X)] &= \psi(\alpha) - \psi(\alpha+\beta) \\
\mathbb{V}[\ln(X)] &= \psi^{(1)}(\alpha) - \psi^{(1)}(\alpha+\beta).
\end{aligned} \tag{A.6}$$

Here $\psi(x) = \frac{d}{dx}\ln(\Gamma(x))$ is the digamma function and $\psi^{(1)}(x) = \frac{d}{dx}\psi(x)$ is the trigamma function. These moments will come in handy when developing the point estimators.

The beta distribution is the conjugate prior to the binomial. This fact makes the beta distribution a natural choice of prior for the win probabilities P. If the probability i beats j is p then the probability j betas i is $1-p$. Since the start and endpoint of each edge was chosen arbitrarily the prior distribution for P should be symmetric about $p = 1/2$. The beta distribution is symmetric if $\alpha = \beta$. Therefore, we assume that the win probabilities are sampled i.i.d. from a symmetric beta distribution:

$$P_k \sim \text{Beta}(\gamma, \gamma). \tag{A.7}$$

Then the posterior distribution for P_k given W_k is:

$$\begin{aligned}
\pi_{P_k}(p) &\propto \binom{n_k}{w_k} p_k^w (1-p)^{n_k - w_k} \frac{p^{\gamma-1}(1-p)^{\gamma-1}}{B(\gamma,\gamma)} \\
&\propto p^{(w_k+\gamma-1)}(1-p)^{(n_k-w_k+\gamma-1)}.
\end{aligned} \tag{A.8}$$

The posterior is supported on $[0,1]$ and has the same functional dependence on p as a

beta distribution hence P_k given w_k is beta distributed:

$$P_k \text{ given } w_k \sim \text{Beta}(w_k + \gamma, n_k - w_k + \gamma). \tag{A.9}$$

So, given n_k, w_k and prior parameter γ:

$$\begin{aligned} &\textbf{Likelihood: } W_k \text{ given } p \sim \text{binomial}(n_k, p) \\ &\textbf{Prior: } P_k \sim \text{Beta}(\gamma, \gamma) \\ &\textbf{Posterior: } P_k \text{ given } w_k \sim \text{Beta}(w_k + \gamma, n_k - w_k + \gamma) \end{aligned} \tag{A.10}$$

Since P_k is beta distributed given the observed wins w_k the moments of the posterior are easy to evaluate (see Equation (A.6)):

$$\begin{aligned} \mathbb{E}[P_k|w_k] &= \frac{w_k + \gamma}{n_k + 2\gamma} \\ \mathbb{V}[P_k|w_k] &= \frac{(w_k + \gamma)(n_k - w_k + \gamma)}{(n_k + 2\gamma)^2(n_k + 2\gamma + 1)} \\ \mathbb{E}[\ln(P_k)|w_k] &= \psi(w_k + \gamma) - \psi(n_k + 2\gamma) \\ \mathbb{V}[\ln(P_k)|w_k] &= \psi^{(1)}(w_k + \gamma) - \psi^{(1)}(n_k + 2\gamma). \end{aligned} \tag{A.11}$$

Equation (A.11) grants a natural interpretation of the prior parameter, γ. The parameter, γ, is, in effect, a fictitious number of games added to the win and loss record of each team. For example, the expected value of the probability $i(k)$ betas $j(k)$ is equal to the percent of games $i(k)$ won against $j(k)$ if $i(k)$ won $w_k + \gamma$ games and lost $n_k - w_k + \gamma$ games.

A.2.1 Choice of Prior Parameter

The prior parameter, γ, can be interpreted as a fictitious number of wins and losses added to the record of each team. In general $\gamma \geq 0$. There are some conventional choices of γ.

These are:

1. Haldane: $\gamma = 0$. The prior distribution is $1/2\delta(x) + 1/2\delta(x-1)$.

2. Jefferys: $\gamma = 1/2$. The prior distribution is proportional to $1/\sqrt{p(1-p)}$.

3. Bayes: $\gamma = 1$. The prior is uniform.

Alternatively, we can fit for γ based on win records. Since we assumed that the win probabilities on each edge were sampled i.i.d. from a gamma distribution, and then the wins were binomially distributed according to the number of events and sampled win probability, the number of wins on each edge is beta-binomial distributed given γ. This means that:

$$W_k \sim \text{Beta-binomial}(\gamma, n_k) \text{ so } \Pr\{W_k = w\} = \binom{n_k}{w} \frac{B(w+\gamma, n-w+\gamma)}{B(\gamma,\gamma)}. \quad \text{(A.12)}$$

Thus, if no prior is assumed on γ:

$$\pi_\gamma(\gamma|n,w) \propto \prod_k \binom{n_k}{w} \frac{B(w_k+\gamma, n_k-w_k+\gamma)}{B(\gamma,\gamma)}. \quad \text{(A.13)}$$

Here the range of the product is not specified as it is implied that the product is taken over all data available. This may include more observed wins than are included in the tournament under study, or the observed win records used to estimate γ may be entirely distinct from the tournament of interest.

It follows from Equation (A.13) that the negative log-likelihood of γ given win records w is (up to an additive constant):

$$\mathcal{G}_\gamma(\gamma|n,w) = \sum_k \ln\left(B(\gamma,\gamma)\right) - \ln\left(B(w_k+\gamma, n_k - w_k\gamma)\right) \quad \text{(A.14)}$$

Here $\mathcal{G}$ is used since the negative log-likelihood is the “Gibbs energy" function.

Substituting in for $B(\alpha, \beta)$ gives:

$$\begin{aligned}\mathcal{G}_\gamma(\gamma|n,w) = \sum_k & \ln\left(\Gamma(\gamma)\right)^2 - \ln\left(\Gamma(2\gamma)\right) \\ & - \ln\left(\Gamma(w_k+\gamma)\right) - \ln\left(\Gamma(n_k - w_k + \gamma)\right) + \ln\left(\Gamma(n_k+2\gamma)\right).\end{aligned} \tag{A.15}$$

Thus γ can be estimated by solving for γ_{MLE}:

$$\gamma_{MLE}(n,w) = \text{argmin}_{\gamma>0}\{\mathcal{G}_\gamma(\gamma|n,w)\} \tag{A.16}$$

The Gibbs energy is fairly cheap to evaluate, γ_{MLE} only needs to be solved for once, and this is a one-dimensional minimization problem, so it is not prohibitively expensive to start by evaluating γ at a series of different sample values. In practice we have found that using γ sampled at geometrically spaced intervals between 10^{-3} and 10^3 gives a good initial sampling of the energy function. The sample with the smallest value of the energy function can then be used to initialize a numerical optimizer.

Initializing in this way avoids the following stability issues. The Gibbs energy function is close to constant for large γ, and is very steep as γ converges to zero. As a result, if initialized too close to zero the first step of the optimizer can easily overshoot the minimum. If initialized at too large a γ the slope of the Gibbs function is so close to zero that the optimizer stops iterating.

A.3 Posterior for Edge Flow

Given a set of win probabilities p the log-odds or logit edge flow is given by:

$$f_k = \text{logit}(p_k) = \ln\left(\frac{p_k}{1-p_k}\right). \tag{A.17}$$

The win probabilities can be recovered from the log-odds via the logistic function:

$$p_k = \text{logistic}(f_k) = \frac{1}{1 + \exp(-f_k)}. \tag{A.18}$$

Note that $\text{logistic}(-f_k) = 1 - p_k$ and $\text{logit}(1 - p_k) = -f_k$.

The HHD is a decomposition of the log-odds edge flow. Therefore, to estimate the components of the HHD we need to be able to estimate the edge-flow. This requires pushing the posterior distribution of the win probabilities forward to the posterior distribution for the edge flow.

Using the standard change of variable formula for probability densities:

$$\pi_F(f) = \pi_P(p(f)|n, w, \gamma)|\frac{d}{df}p(f)| \tag{A.19}$$

where $p(f) = \text{logistic}(f)$ and $\pi_P(p|w)$ is given by the beta distribution with parameters $w + \gamma, n - w + \gamma$ (see Equation A.10). Thus:

$$\begin{aligned} \pi_F(f) &= \text{Beta}(\text{logistic}(f)|n, w, \gamma) \left|\frac{d}{df}\text{logistic}(f)\right| \\ &= \frac{\text{logistic}(f)^{w+\gamma-1}\text{logistic}(-f)^{w+\gamma-1}}{B(w + \gamma, n - w + \gamma)} \left|\frac{d}{df}\text{logistic}(f)\right| \end{aligned} \tag{A.20}$$

The derivative is:

$$\begin{aligned} \frac{d}{df}\text{logistic}(f) &= \frac{d}{df}\left(1 + \exp(-f)\right)^{-1} = \frac{\exp(-f)}{(1 + \exp(-f))^2} \\ &= \frac{\exp(-f)}{1 + 2\exp(-f) + \exp(-2f)} = \frac{-1}{\exp(f) + 2 + \exp(-f)} \\ &= \frac{1}{(1 + \exp(f))(1 + \exp(-f))} = \text{logistic}(f)\text{logistic}(-f). \end{aligned} \tag{A.21}$$

Therefore the posterior for the log-odds/logit edge-flow is:

$$.\pi_F(f|n,w,\gamma) = \frac{1}{B(w+\gamma, n-w+\gamma)}\text{logistic}(f)^{w+\gamma}\text{logistic}(-f)^{w+\gamma}. \quad \text{(A.22)}$$

A.4 Point Estimators for Edge Flow

Given the posterior for the win probabilities and the edge flow we can compute point estimators for the edge flow. We will consider two point estimators, the MAP estimator, and the conditional expectation.

A.4.1 The MAP Estimator

To compute the MAP estimator we need to maximize the posterior distribution of the edge flow. This is equivalent to minimizing the negative log of the posterior:

$$\mathcal{G}_F(f|n,w,\gamma) = (w+\gamma)\ln\left(\text{logistic}(f)\right) + (n-w+\gamma)\ln\left(\text{logistic}(-f)\right). \quad \text{(A.23)}$$

Substituting in for the logistic function gives:

$$\mathcal{G}_F(f|n,w,\gamma) = -(w+\gamma)\ln\left(1+\exp(-f)\right) - (n-w+\gamma)\ln\left(1+\exp(f)\right). \quad \text{(A.24)}$$

Then, differentiating with respect to f:

$$\begin{aligned}\frac{d}{df}\mathcal{G}_F(f|n,w,\gamma) &= -(w+\gamma)\frac{-\exp(-f)}{1+\exp(-f)} - (n-w+\gamma)\frac{\exp(f)}{1+\exp(f)} \\ &= (w+\gamma)\frac{1}{1+\exp(f)} - (n-w+\gamma)\frac{1}{1+\exp(-f)} \\ &= (w+\gamma)\text{logistic}(-f) - (n-w+\gamma)\text{logistic}(f) \\ &= (w+\gamma)(1-\text{logistic}(f)) - (n-w+\gamma)\text{logistic}(f) \\ &= (w+\gamma) - (n+2\gamma)\text{logistic}(f).\end{aligned} \tag{A.25}$$

Then, setting to zero:

$$(n+2\gamma)\text{logistic}(f) = w+\gamma \tag{A.26}$$

Let $f = \text{logit}(p)$. Then $\text{logistic} f = p$ and:

$$p = \frac{w+\gamma}{n+2\gamma}. \tag{A.27}$$

Since this is the only solution to the equations the negative log-likelihood has one extrema. Since the posterior has decaying tails and the negative log-likelihood is convex (see Section A.5) this extrema must be a minimizer of the negative log-likelihood. Therefore the MAP estimator for f is the logit of the expected win probability given the data, and is equal to the logit of the win frequency after adding γ fictitious wins and losses to the record.

$$f_{\text{MAP}}(n,w,\gamma) = \text{logit}(\mathbb{E}[P|n,w,\gamma]) = \ln\left(\frac{w+\gamma}{n-w+\gamma}\right). \tag{A.28}$$

A.4.2 Conditional Expectation

The conditional expectation of the log-odds given the data and prior is:

$$f_{\text{exp}}(n, w, \gamma) = \mathbb{E}[\text{logit}(P)|w] = \mathbb{E}[\ln(P)|w] - \mathbb{E}[\ln(1-P)|w] \tag{A.29}$$

We know that the win probabilities P, and loss probabilities $1 - P$, are beta distributed when conditioned on the data with parameters $w + \gamma, n - w + \gamma$ and $n - w + \gamma, w + \gamma$ respectively. The expectation of the log of a beta distributed random variable is given by the digamma function (see Equation (A.6)). Thus:

$$\begin{aligned} f_{\text{exp}}(n, w, \gamma) &= (\psi(w+\gamma) - \psi((w+\gamma) + (n-w+\gamma))) \\ &\quad - (\psi(n-w+\gamma) - \psi((n-w+\gamma) + (w+\gamma))) \\ &= \psi(w+\gamma) - \psi(n+2\gamma) - \psi(n-w+\gamma) + \psi(n-w+\gamma). \end{aligned} \tag{A.30}$$

Therefore the conditional expectation of F given the data is:

$$f_{\text{exp}}(n, w, \gamma) = \mathbb{E}[F|n, w, \gamma] = \psi(w+\gamma) - \psi(n-w+\gamma). \tag{A.31}$$

The digamma function $\psi(x) = \frac{d}{dx}\ln(\Gamma(x))$ is the logarithmic derivative of the gamma function. The digamma function satisfies the recursion:

$$\psi(x+1) = \psi(x) + \frac{1}{x}. \tag{A.32}$$

This gives the following recursion for the conditional expectation of the log-odds that can be updated live each time a win or loss is observed:

Let h index the game observed. Let w_h be the number of wins observed in the first h

games. Initialize $f_{exp}(h, w_h, \gamma) = 0$. Then:

$$\begin{aligned} &\text{if win: } f_{exp}(h+1, w_h+1, \gamma) = f_{exp}(h, w_h, \gamma) + \frac{1}{w_h + \gamma} \\ &\text{if lose: } f_{exp}(h+1, w_h, \gamma) = f_{exp}(h, w_h, \gamma) - \frac{1}{n - w_h + \gamma}. \end{aligned} \tag{A.33}$$

That is, for every win we increase the conditional expectation by one over the number of previously observed wins (including fictitious wins), and for every loss we decrease the conditional expectation by one over the number of previously observed losses (including fictitious losses). Notice that events that are surprising (events we have not seen yet, or have seen infrequently) lead to smaller corrections than events that are not surprising.

A.4.3 Comparison

The digamma function has asymptotic expansion:

$$\psi(x) \sim \ln(x) - \frac{1}{2}x + \mathcal{O}(x^{-2}) \tag{A.34}$$

and is bounded by:

$$\psi(x) \in \ln(x) - \left[\frac{1}{x}, \frac{1}{2x}\right]. \tag{A.35}$$

Thus:

$$\begin{aligned} f_{\text{exp}}(n, w, \gamma) \sim &\ln(w + \gamma) - \ln(n - w + \gamma) - \frac{1}{2(w + \gamma)} + \frac{1}{2(n - w + \gamma)} \\ &+ \mathcal{O}((w + \gamma)^{-2}) + \mathcal{O}((n - w + \gamma)^{-2}). \end{aligned} \tag{A.36}$$

Simplifying:

$$\begin{aligned} f_{\text{exp}}(n,w,\gamma) &\sim \ln\left(\frac{w+\gamma}{n-w+\gamma}\right) + \frac{1}{2}\frac{(w+\gamma)-(n-w+\gamma)}{(w+\gamma)(n-w+\gamma)} \\ &\quad + \mathcal{O}((w+\gamma)^{-2}) + \mathcal{O}((n-w+\gamma)^{-2}) \\ &= \ln\left(\frac{w+\gamma}{n-w+\gamma}\right) + \frac{1}{2}\frac{2w-n}{(w+\gamma)(n-w+\gamma)} \\ &\quad + \mathcal{O}((w+\gamma)^{-2}) + \mathcal{O}((n-w+\gamma)^{-2}). \end{aligned} \tag{A.37}$$

Notice that the logarithmic term is the MAP estimator for the log-odds edge flow. Thus the conditional expectation differs from the MAP estimator by the bias $(2w-n)/((w+\gamma)(n-w+\gamma))$, and additional order $(w+\gamma)^{-2}, (n-w+\gamma)^{-2}$ terms. This bias will appear throughout the subsequent analysis. Since the bias is order $(w+\gamma)^{-1}, (n-w+\gamma)^{-1}$, and both the number of wins and losses are order n in expectation, the conditional expectation is expected to converge to the MAP estimator as more events are observed.

By using the bounds on the digamma function we can bound the difference in the estimators:

$$\begin{aligned} f_{exp}(n,w,\gamma) &\leq \ln\left(\frac{w+\gamma}{n-w+\gamma}\right) - \frac{1}{2(w+\gamma)} + \frac{1}{(n-w+\gamma)} \\ f_{exp}(n,w,\gamma) &\geq \ln\left(\frac{w+\gamma}{n-w+\gamma}\right) - \frac{1}{(w+\gamma)} + \frac{1}{2(n-w+\gamma)} \end{aligned} \tag{A.38}$$

Therefore the conditional expectation for the log-odds edge flow converges to the MAP estimator with difference:

$$\begin{aligned} f_{\text{exp}}(n,w,\gamma) - f_{\text{MAP}}(n,w,\gamma) =& \frac{1}{2}\frac{2w-n}{(w+\gamma)(n-w+\gamma)} \\ &+ \mathcal{O}((w+\gamma)^{-2}) + \mathcal{O}((n-w+\gamma)^{-2}). \end{aligned} \tag{A.39}$$

Moreover, the difference in the estimators is bounded by:

$$f_{\text{exp}}(n, w, \gamma) - f_{\text{MAP}}(n, w, \gamma) \in \frac{1}{2} \frac{2w - n}{(w + \gamma)(n - w + \gamma)} + \left[\frac{-1}{2(w + \gamma)}, \frac{1}{2(n - w + \gamma)} \right]. \tag{A.40}$$

Therefore the conditional expectation for the log-odds edge flow converges to the MAP estimator with convergence rate $\mathcal{O}((w)^{-1}) + \mathcal{O}((n - w)^{-1})$.

A.5 Properties of the Posterior

A.5.1 Tail Behavior

The tails of the posterior distribution of F decay exponentially. As f converges to infinity logistic(f) converges to one and logistic$(-f)$ converges to $\exp(-f)$. As f converges to infinity $\cosh(f/2)^2$ converges to $\exp(f)$. Therefore:

$$\begin{aligned} \lim_{f \to \infty} \pi_F(f | n, w, \gamma) &\propto \lim_{f \to \infty} \exp(-(n - w + \gamma) f) \\ \lim_{f \to \infty} \pi_F(f | n, w, \gamma) &\propto \lim_{f \to \infty} \exp(-(w + \gamma) f) \end{aligned} \tag{A.41}$$

where the second equation follows by symmetry.

Thus the tails of posterior distribution of F decay exponentially with rates equal to the number of observed losses plus fictitious losses, $(n - w + \gamma)$, as f goes to infinity, and equal to the number of observed wins plus fictitious wins, $(w + \gamma)$, as f goes to negative infinity. Therefore each observed loss controls the upper estimate of the log-odds, and each observed win controls the lower estimate of the log-odds. It also follows that, as long as $\gamma > 0$ the distribution is well defined since it must integrate to a constant.

A.5.2 Variance

The variance in the posterior for the edge flow, like the mean of the posterior, can be evaluated using known moments of the beta distribution (see Equation (A.6)). The variance in the posterior is:

$$\begin{aligned}\mathbb{V}[F|n,w,\gamma] &= \mathbb{V}[\ln(P) - \ln(1-P)|n,w,\gamma] \\ &= \mathbb{V}[\ln(P)|n,w,\gamma] - 2\text{Cov}[\ln(P),\ln(1-P)|n,w,\gamma] \\ &+ \mathbb{V}[\ln(1-P)|n,w,\gamma].\end{aligned} \tag{A.42}$$

If X is beta distributed then the variance in the log of X is $\psi^{(1)}(\alpha)+\psi^{(1)}(\alpha+\beta)$ where $\psi^{(1)}(x)$ is the trigamma function (see Equation (A.6)). The covariance is also known:

$$\text{Cov}[\ln(X),\ln(1-X)] = \psi^{(1)}(\alpha+\beta). \tag{A.43}$$

Therefore:

$$\begin{aligned}\mathbb{V}[F|n,w,\gamma] =& \psi^{(1)}(w+\gamma) + \psi^{(1)}(n+2\gamma) - 2\psi^{(1)}(n+2\gamma) \\ &+ \psi^{(1)}(n-w+\gamma) + \psi^{(1)}(n+2\gamma).\end{aligned} \tag{A.44}$$

Cancelling the repeated terms gives:

$$\mathbb{V}[F|n,w,\gamma] = \psi^{(1)}(w+\gamma) + \psi^{(1)}(n-w+\gamma). \tag{A.45}$$

Differentiating the asymptotic expansion of the digamma function gives an asymptotic

expansion for the trigamma function. This yeilds the approximation:

$$\mathbb{V}[F|n,w,\gamma] \sim \frac{1}{w+\gamma} + \frac{1}{n-w+\gamma} + \mathcal{O}((w+\gamma)^{-2}) + \mathcal{O}((n-w+\gamma)^{-2}). \quad \text{(A.46)}$$

A.5.3 Convexity

The derivative of the negative log posterior with respect to f was (see Equation (A.25)):

$$-\frac{d}{df}\mathcal{G}_F(f|n,w,\gamma) = -(w+\gamma) + (n+2\gamma)\text{logistic}(f). \quad \text{(A.47)}$$

Therefore, the second derivative is:

$$\frac{d^2}{df^2}\mathcal{G}_F(f|n,w,\gamma) = (n+2\gamma)\frac{\exp(-f)}{(1+\exp(-f))^2}. \quad \text{(A.48)}$$

The logistic function is monotonically increasing since $\frac{\exp(-f)}{(1+\exp(-f))^2} > 0$. Therefore:

$$-\frac{d^2}{df^2}\mathcal{G}_F(f|n,w,\gamma) > 0 \quad \text{(A.49)}$$

It follows that the negative log of the posterior is convex.

A.6 Sample Size Requirements

With the variance and tail behavior of the posterior in hand we can establish some simple sample size requirements. The tails of the posterior decay exponentially with rate given by the number of observed wins (plus fictitious wins), and the number of observed losses (plus fictitious losses). Similarly, the variance converges to one over the number of observed wins (plus fictitious wins) plus one over the number of observed losses (plus fictitious losses).

Thus to ensure that the tails decay quickly, and that the variance is sufficiently small, *both* the number of observed wins and the number of observed losses must be large.

Suppose that we set an upper bound, ϵ^2, on an acceptable variance in the posterior. Then on every edge k we require that:

$$\psi^{(1)}(w_k + \gamma) + \psi^{(1)}(n_k - w_k + \gamma) \leq \epsilon^2. \tag{A.50}$$

This condition can be easily checked given a win record. In essence it requires that $\min\{w_k + \gamma, n_k - w_k + \gamma\} \geq 2/\epsilon^2$. Note that both the observed wins and losses must be large to satisfy this requirement.

If $P \sim \text{Beta}(\gamma, \gamma)$ then the expected variance on each edge after observing n samples is $\mathbb{E}[\psi^{(1)}(W_k+\gamma)+\psi^{(1)}(n_k-W_k+\gamma)]$ where W_k is beta-binomial distributed with parameters n_k and γ. Since $\text{Beta}(\gamma, \gamma)$ is symmetric $\mathbb{E}[\psi^{(1)}(W_k + \gamma)] = \mathbb{E}[\psi^{(1)}(n_k - W_k + \gamma)]$ so $\mathbb{E}[\psi^{(1)}(W_k + \gamma) + \psi^{(1)}(n_k - W_k + \gamma)]$ equals $2\mathbb{E}[\psi^{(1)}(W_k + \gamma)]$. Therefore, a lower bound on the sample size required is given by solving for n such that:

$$2\sum_{w=0}^{n} \begin{pmatrix} n \\ j \end{pmatrix} \frac{B(w + \gamma, n - w + \gamma)}{B(\gamma, \gamma)} \psi^{(1)}(w + \gamma) < \epsilon^2. \tag{A.51}$$

Thus, the minimum sample size needed to ensure that the expected variance in the posterior is less than ϵ^2 is:

$$n_{\text{min}} = \min \left\{ n \text{ s. t. } \sum_{w=0}^{n} \begin{pmatrix} n \\ w \end{pmatrix} \frac{B(w + \gamma, n - w + \gamma)}{B(\gamma, \gamma)} \psi^{(1)}(w + \gamma) < \frac{1}{2}\epsilon^2 \right\}. \tag{A.52}$$

This sets a minimum sample size we expect to need before observing the win record (assuming γ is known from prior data).

In practice we want the expected size of the uncertainty in the posterior on an edge to be small relative to expected size of the point estimators on the edge. This leads to minimum sample size requirements of the form:

$$n_{\text{min}} = \min \left\{ n \text{ s. t. } \sum_{w=0}^{n} \left(\begin{array}{c} n \\ w \end{array} \right) \frac{B(w+\gamma, n-w+\gamma)}{B(\gamma,\gamma)} \frac{|f_{exp}(n,w,\gamma)|}{\sqrt{\mathbb{V}[F|n,w,\gamma]}} > \text{SNR} \right\}. \tag{A.53}$$

This condition ensures that the expected signal to noise ratio (conditional expectation divided by standard devitation) is greater than a desired SNR. Numerical tests show that the minimum sample size scales in the desired SNR squared, and grows linearly in γ. The required sample size increase with increasing γ despite the fact that increasing gamma decreases the variance since increasing γ also decreases the average size of the signal (conditional expectation).

A.7 Asymptotic Expansion of Expectation

Suppose that the number of observed events is large. Then the distribution of W/n becomes increasingly tightly concentrated around its mean (variance order $n^{-1/2}$). Our estimators can be expressed as smooth functions of W/n. Therefore, if the true f was known the expected value of our estimators could be approximated analytically in the large n limit. Then by comparing the expected value of the estimators to the true f we can identify sources of bias in the estimation, and how quickly they converge to zero as the number of events grows. Similar analysis can be used to compute the variance in the estimators given a true log-odds f and number of events.

These analyses rely on the following analytical method for the asymptotic expansion of expectation.

Let $X(n)$ be a random variable which takes on values in $\mathcal{R}^d$ and n is a sample size. Suppose that the h^{th} central moments of X are $\mathcal{O}(n^{-\lceil h/2 \rceil})$. Then the variance is order -1 in n, so the distribution is concentrating about its mean. Let $\bar{x}(n) = \mathbb{E}[X(n)]$. We will usually assume this is independent of n so will omit the dependence on $\bar{x}$ unless necessary.

Let $g(X)$ be a continuously differentiable function with a convergent Taylor series in an open neighborhood containing $\bar{x}$.

Let α be a multi-index $\alpha = \alpha_1, \alpha_2, ...\alpha_d$. Let $|\alpha| = \sum_{j=1}^{d} \alpha_j$ be the magnitude of the multi-index. Let x^α denote $\prod_j x_j^{\alpha_j}$. Then let $M^\alpha[X] = \mathbb{E}[(X - \bar{x})^\alpha]$ denote the α central moment, and $M^h[X]$ denote the tensor consisting of all $|\alpha| = h$ order moments. Similarly let $\partial^\alpha g(x) = \partial_{x_1}^{\alpha_1} ... \partial_{x_d}^{\alpha_d} g(x)$, and $\partial^h g(x)$ be the tensor containing all $|\alpha| = h$ order partial deriviatives of $g(x)$.

Then, Taylor expanding $g(x)$ about $\bar{x}$ gives:

$$g(x) = \sum_{h=0}^{\infty} \frac{1}{h!} \langle \partial^h g(\bar{x}), (x - \bar{x})^h \rangle \tag{A.54}$$

for x inside the radius of convergence of the power series. Here $\langle A, B \rangle$ denotes the tensor inner product:

$$\langle \partial^h g(\bar{x}), (x - \bar{x})^h \rangle = \sum_{\alpha || \alpha| = h} \partial^\alpha g(\bar{x}) (x - \bar{x})^\alpha \tag{A.55}$$

Therefore, the expected value of $g(X(n))$ is:

$$\mathbb{E}[g(X(n))] = \sum_{h=0}^{\infty} \frac{1}{h!} \langle \partial^h g(\bar{x}), M^h[X(n)] \rangle \tag{A.56}$$

This gives the approximation:

$$\mathbb{E}[g(X(n))] = g(\bar{x}) + \frac{1}{2} \langle \partial^2 g(\bar{x}), \mathbb{V}[X(n)] \rangle + \mathcal{O}(n^{-2}). \tag{A.57}$$

In this equation $\partial^2 g(\bar{x})$ is the Hessian of $g(x)$ at $\bar{x}$. Therefore, all that is needed to apply this approximation is the expected value of $X(n)$, the covariance matrix of $X(n)$, and the Hessian of $g(x)$ at $\bar{x}$.

The same framework can be used to approximate the variance in $g(X(n))$. Squaring the Taylor expansion of $g(x)$ gives:

$$\begin{aligned} g(x)^2 = g(\bar{x})^2 + 2g(\bar{x})\langle \partial g(\bar{x}), (x-\bar{x})\rangle + \langle \partial g(\bar{x}), (x-\bar{x})\rangle^2 \\ + 2g(\bar{x})\langle \partial^2 g(\bar{x}), (x-\bar{x})^2\rangle + \mathcal{O}((x-\bar{x})^3) \end{aligned} \tag{A.58}$$

Therefore:

$$\begin{aligned} \mathbb{E}[g(X(n))^2] = g(\bar{x})^2 + \mathbb{E}\left[\langle \partial g(\bar{x}), (X(n)-\bar{x})\rangle^2\right] \\ + 2g(\bar{x})\langle \partial^2 g(\bar{x}), \mathbb{V}[X(n)]\rangle + \mathcal{O}(n^{-2}). \end{aligned} \tag{A.59}$$

Then, subtracting off the asymptotic approximation to $\mathbb{E}[g(X(n))]^2$ leaves:

$$\mathbb{V}[g(X(n))] = \mathbb{E}\left[\langle \partial g(\bar{x}), (X(n)-\bar{x})\rangle^2\right] + \mathcal{O}(n^{-2}). \tag{A.60}$$

This can be written more cleanly by letting $\nabla g(x) = \partial g(\bar{x})$ be the gradient. Then:

$$\begin{aligned} \mathbb{V}[g(X(n))] &= \langle \nabla g(\bar{x}) \nabla g(\bar{x})^T, \mathbb{V}[X(n)]\rangle + \mathcal{O}(n^{-2}) \\ &= \nabla g(\bar{x})^T \mathbb{V}[X(n)] \nabla g(\bar{x}) + \mathcal{O}(n^{-2}). \end{aligned} \tag{A.61}$$

Thus, if $X(n)$ has central moments order $n^{-\lceil h/2 \rceil}$, $\mathbb{E}[X(n)] = \bar{x}$, and $g(x)$ is analytic

about $\bar{x}$ then:

$$\begin{aligned}\mathbb{E}[g(X(n))] &= g(\bar{x}) + \frac{1}{2}\langle \partial^2 g(\bar{x}), \mathbb{V}[X(n)]\rangle + \mathcal{O}(n^{-2}) \\ \mathbb{V}[g(X(n))] &= \langle \nabla g(\bar{x}) \nabla g(\bar{x})^T, \mathbb{V}[X(n)]\rangle + \mathcal{O}(n^{-2}) \\ &= \nabla g(\bar{x})^T \mathbb{V}[X(n)] \nabla g(\bar{x}) + \mathcal{O}(n^{-2}).\end{aligned} \tag{A.62}$$

A.8 Asymptotic Bias and Uncertainty in Point Estimators

Given a true f what is the expected value of the point estimators given a sample of n events? What is the variance in the point estimators given a sample?

To compute the expected value and the variance we will use Equation (A.62) where $X(n) = W/n$ and $g(X)$ is a point estimator (either the MAP estimator A.28 or the conditional expectation A.31).

Fix f and set $p = \text{logistic}(f)$. Then W is binomially distributed so $X(n)$ has mean p, variance $p(1-p)/n$ and higher order central moments that are order n^{-2} or higher. In terms of X the estimators are:

$$\begin{aligned} f_{\text{MAP}}(n, nX(n), \gamma) &= \ln\left(\frac{X(n) + \gamma/n}{1 - X(n) + \gamma/n}\right) \\ f_{\text{exp}}(n, nX(n), \gamma) &= \psi(n(X(n) + \gamma/n)) - \psi(n(1 - X(n) + \gamma/n)). \end{aligned} \tag{A.63}$$

A.8.1 MAP Estimator

First let's find the expected value of the MAP estimator for f given f. This will require differentiating $f_{\text{MAP}}(n, nx, \gamma)$ in x:

$$\begin{aligned}\partial_x f_{\text{MAP}}(n, nx, \gamma) &= \frac{n}{n(x + \gamma/n)} - \frac{-n}{n(1 - x + \gamma/n)} = \frac{1}{x + \gamma/n} + \frac{1}{1 - x + \gamma/n} \\ &= \frac{1 + 2\gamma/n}{(x + \gamma/n)(1 - x + \gamma/n)}.\end{aligned} \tag{A.64}$$

The second derivative is:

$$\begin{aligned}
\partial_x^2 f_{\text{MAP}}(n, nx, \gamma) &= \frac{-1}{(x+\gamma/n)^2} + \frac{1}{(1-x+\gamma/n)^2} \\
&= -\frac{(1-x+\gamma/n)^2 - (x+\gamma/n)^2}{(x+\gamma/n)^2(1-x+\gamma/n)^2} \\
&= -\frac{((1+2\gamma/n) - (x+\gamma/n))^2 - (x+\gamma/n)^2}{(x+\gamma/n)^2(1-x+\gamma/n)^2} \\
&= -\frac{(1+2\gamma/n)^2 - 2(1+2\gamma/n)(x+\gamma/n)}{(x+\gamma/n)^2(1-x+\gamma/n)^2} \\
&= -\frac{(1+2\gamma/n)((1+2\gamma/n) - 2(x+\gamma/n))}{(x+\gamma/n)^2(1-x+\gamma/n)^2} \\
&= -\frac{(1+2\gamma/n)(1-2x)}{(x+\gamma/n)^2(1-x+\gamma/n)^2}
\end{aligned} \tag{A.65}$$

Therefore, using equation A.62:

$$\begin{aligned}
\mathbb{E}[f_{\text{MAP}}(n, W, \gamma)|p] &= f_{\text{MAP}}(n, np, \gamma) + \frac{1}{2}\frac{(1+2\gamma/n)(1-2p)}{(p+\gamma/n)^2(1-p+\gamma/n)^2}\frac{p(1-p)}{n} + \mathcal{O}(n^{-2}) \\
&= \ln\left(\frac{p+\gamma/n}{1-p+\gamma/n}\right) - \frac{1}{2}\frac{(1+2\gamma/n)(1-2p)}{(p+\gamma/n)^2(1-p+\gamma/n)^2}\frac{p(1-p)}{n} + \mathcal{O}(n^{-2}).
\end{aligned} \tag{A.66}$$

Since we only desire an order n^{-2} accurate approximation the coefficient appearing in front of the middle term only needs to be approximated to order 1 accuracy in n. Since γ is fixed γ/n vanishes as n gets large. Therefore:

$$\begin{aligned}
\mathbb{E}[f_{\text{MAP}}(n, W, \gamma)|p] &= \ln\left(\frac{p+\gamma/n}{1-p+\gamma/n}\right) - \frac{1}{2}\frac{(1-2p)}{(p)^2(1-p)^2}\frac{p(1-p)}{n} + \mathcal{O}(n^{-2}) \\
&= \ln\left(\frac{p+\gamma/n}{1-p+\gamma/n}\right) - \frac{1}{2}\frac{(1-2p)}{p(1-p)}\frac{1}{n} + \mathcal{O}(n^{-2})
\end{aligned} \tag{A.67}$$

Next, expand the first term in small γ/n. This gives:

$$\begin{aligned}\ln\left(\frac{p+\gamma/n}{1-p+\gamma/n}\right) &= = \ln\left(\frac{p}{1-p}\right) + \frac{1}{p}\frac{\gamma}{n} - \frac{1}{1-p}\frac{\gamma}{n} + \mathcal{O}(n^{-2}) \\ &= \ln\left(\frac{p}{1-p}\right) + \frac{1-2p}{p(1-p)}\frac{\gamma}{n} + \mathcal{O}(n^{-2}).\end{aligned} \tag{A.68}$$

Therefore:

$$\begin{aligned}\mathbb{E}[f_{\text{MAP}}(n,W,\gamma)|p] &= \ln\left(\frac{p}{1-p}\right) + \frac{(1-2p)}{p(1-p)}\frac{\gamma}{n} - \frac{1}{2}\frac{(1-2p)}{(p)^2(1-p)^2}\frac{p(1-p)}{n} + \mathcal{O}(n^{-2}) \\ &= \ln\left(\frac{p}{1-p}\right) + \frac{(1-2p)}{p(1-p)}\frac{(\gamma-1/2)}{n} + \mathcal{O}(n^{-2}) \\ &= \text{logit}(p) - \frac{(2p-1)}{p(1-p)}\frac{(\gamma-1/2)}{n} + \mathcal{O}(n^{-2}).\end{aligned} \tag{A.69}$$

Let's express this in terms of f. First, $\text{logit}(p) = f$ so the first term is the true log-odds. This leaves the second term (order n^{-1} term). Substituting in $p = \text{logistic}(f)$ gives:

$$\begin{aligned}\frac{1}{p(1-p)} &= (1+\exp(-f))(1+\exp(f)) = \exp(f) + 2 + \exp(-f) \\ &= 2(1+\cosh(f/2)) = (2\cosh(f/2))^2 \\ (2p-1) &= \frac{2}{1+\exp(-f)} - 1 = \frac{1-\exp(-f)}{1+\exp(-f)} \\ &= \frac{\exp(f/2)-\exp(-f/2)}{\exp(f/2)+\exp(-f/2)} = \tanh(f/2) \\ \frac{2p-1}{p(1-p)} &= (2\cosh(f/2))^2\frac{\sinh(f/2)}{\cosh(f/2)} = 4\cosh(f/2)\sinh(f/2) = 2\sinh(f).\end{aligned} \tag{A.70}$$

Therefore, in the limit of large n the expected value of the MAP estimator for the log-odds given true log-odds f is:

$$\mathbb{E}[f_{\text{MAP}}(n,W,\gamma)|f] = f - 2\sinh(f)\frac{(\gamma-1/2)}{n} + \mathcal{O}(n^{-2}) \tag{A.71}$$

Here we can clearly see that the MAP estimator has two primary sources of bias. The first, $-2\sinh(f)\gamma/n$ is the bias due to the prior. The prior assumes that it is more likely to sample small f than large f, hence the MAP estimator errs on the side of underestimating f. Since $\sinh(f)$ is positive when f is positive and negative when f is negative this bias makes f smaller in magnitude. The larger γ the tighter the prior is distributed about $f = 0$, so the stronger the bias introduced. This bias decreases order n^{-1} as the more events are observed the more the MAP estimate is informed by the observed data than by prior expectation.

The second source of bias comes from the fact that W/n has nonzero variance, and f_{MAP} is a nonlinear function of W/n that is concave up when $W/n > 1/2$ and concave down when $W/n < 1/2$. Thus, sampling error leads to a systematic overestimate of the magnitude of f when using the MAP estimate.

These two biases balance out when $\gamma = 1/2$ (Jefferys' prior). Then any bias in the MAP estimator is $\mathcal{O}(n^{-2})$.

Next we compute the variance in the MAP estimator given f in the limit of large n. To compute this variance we use Equation (A.62). This gives:

$$\mathbb{V}[f_{\text{MAP}}(n, nX(n), \gamma)|p] = \left(\frac{(1 + 2\gamma/n)}{(p + \gamma/n)(1 - p + \gamma/n)}\right)^2 \frac{p(1-p)}{n} + \mathcal{O}(n^{-2}). \quad \text{(A.72)}$$

Again, only keeping terms up to $\mathcal{O}(n^{-1})$:

$$\mathbb{V}[f_{\text{MAP}}(n, nX(n), \gamma)|p] = \frac{1}{p(1-p)}\frac{1}{n} + \mathcal{O}(n^{-2}). \quad \text{(A.73)}$$

Then, substituting in $(p(1-p))^{-1} = (2\cosh(f/2))^2$ gives the following result, in the limit of large n the variance in the MAP estimator for the log-odds given true log-odds f

is:

$$\mathbb{V}[f_{\text{MAP}}(n, W, \gamma)|f] = (2\cosh(f/2))^2 \frac{1}{n} + \mathcal{O}(n^{-2}) \tag{A.74}$$

A.8.2 Conditional Expectation

The same types of calculations can be performed for the conditional expectation of the log-odds. This can be streamlined by recalling that the conditional expectation only differs from the MAP estimate by a term which is $\mathcal{O}(n^{-1})$. In particular:

$$f_{\text{exp}}(n, W, p) = f_{\text{MAP}}(n, W, p) + \frac{1}{2} \frac{2W/n - 1}{(W/n + \gamma/n)(1 - W/n + \gamma/n)} \frac{1}{n} + \mathcal{O}(n^{-2}) \tag{A.75}$$

Since this difference is already $\mathcal{O}(n^{-1})$ we only need to approximate to $\mathcal{O}(n^{-1})$ which is equivalent to approximating the prefix to order 1 in n. This means we simply replace W/n with its expectation, p, and then drop the vanishing γ/n terms. This gives the familiar form:

$$\mathbb{E}[f_{\text{exp}}(n, W, p)|p] = \mathbb{E}[f_{\text{MAP}}(n, W, p)|p] + \frac{1}{2} \frac{2p - 1}{p(1 - p)} \frac{1}{n} + \mathcal{O}(n^{-2}) \tag{A.76}$$

This also implies that, to order $\mathcal{O}(n^{-1})$ the variance in the conditional expectation equals the variance in the MAP estimator. Therefore, in the limit of large n the variance in the MAP estimator for the log-odds given true log-odds f is:

$$\begin{aligned} \mathbb{E}[f_{\text{exp}}(n, W, \gamma)|f] &= f - 2\sinh(f) \frac{(\gamma - 1)}{n} + \mathcal{O}(n^{-2}) \\ \mathbb{V}[f_{\text{exp}}(n, W, \gamma)|f] &= (2\cosh(f/2))^2 \frac{1}{n} + \mathcal{O}(n^{-2}) \end{aligned} \tag{A.77}$$

So, in the limit of large n the only order n^{-1} difference in the expectation of the two point estimators is how strongly the uncertainty in the sampled W/n couples with the

nonlinearity in the estimator. This coupling is slightly larger for the conditional expectation than the MAP estimator, hence the estimate has a larger bias due to sampling uncertainty. The conditional expectation is asympotitically unbiased to order n^{-1} if $\gamma = 1$, that is for the Bayes (uniform) prior.

A.8.3 Asymptotic Correctors and Additional Sample Size Requirements

Can we correct for the biases in the estimators?

The bias in the estimators are:

$$\begin{aligned}\mathbb{E}[f_{\text{MAP}}(n, W, \gamma)|f] - f &= -2\sinh(f)\frac{(\gamma - 1/2)}{n} \\ \mathbb{E}[f_{\text{MAP}}(n, W, \gamma)|f] - f &= -2\sinh(f)\frac{(\gamma - 1)}{n}.\end{aligned} \tag{A.78}$$

These could be computed explicitly if f was known, however the entire motivation for the estimation framework is that f is unknown. However, if we only want to eliminate the bias to order $\mathcal{O}(n^{-1})$ then we only need an order $\mathcal{O}(n^{-1})$ accurate approximation of the bias. Since the bias is a smooth function of f tomes $1/n$ this only requires an order 1 accurate approximation to f. Either of the estimators give an order 1 approximation to f so we can define de-biased estimators:

$$\begin{aligned}f_{\text{MAP}*}(n, W, \gamma) &= f_{\text{MAP}}(n, W, \gamma) + 2\sinh(f_{\text{MAP}}(n, W, \gamma))\frac{(\gamma - 1/2)}{n} \\ f_{\text{exp}*}(n, W, \gamma) &= f_{\text{exp}}(n, W, \gamma) + 2\sinh(f_{\text{exp}}(n, W, \gamma))\frac{(\gamma - 1)}{n}\end{aligned}. \tag{A.79}$$

These estimators will have the same variance (to order n^{-1}) as the original estimators

since the correction is the addition of an order n^{-1} term. Then the debiased estimators:

$$\begin{aligned} f_{\text{MAP}*}(n, W, \gamma) &= f_{\text{MAP}}(n, W, \gamma) + 2\sinh(f_{\text{MAP}}(n, W, \gamma))\frac{(\gamma - 1/2)}{n} \\ f_{\text{exp}*}(n, W, \gamma) &= f_{\text{exp}}(n, W, \gamma) + 2\sinh(f_{\text{exp}}(n, W, \gamma))\frac{(\gamma - 1)}{n} \end{aligned} . \tag{A.80}$$

have identical expectations and variances to order n^{-1} and:

$$\begin{aligned} \mathbb{E}[f_{\text{MAP}*}(n, W, \gamma)|f] &= f + \mathcal{O}(n^{-2}) \\ \mathbb{E}[f_{\text{exp}*}(n, W, \gamma)|f] &= f + \mathcal{O}(n^{-2}) \\ \mathbb{V}[f_{\text{MAP}*}(n, W, \gamma)|f] &= (2\cosh(f/2))^2\frac{1}{n} + \mathcal{O}(n^{-2}) \\ \mathbb{V}[f_{\text{exp}*}(n, W, \gamma)|f] &= (2\cosh(f/2))^2\frac{1}{n} + \mathcal{O}(n^{-2}). \end{aligned} \tag{A.81}$$

Therefore, when n is large we can de-bias the point estimators to order n^{-1}.

Note that this only ensures that the point estimators have the correct expected value to order n^{-2}, and not that they are guaranteed to have the correct value, nor the correct expectation for finite n. Moreover this analysis has only followed the order n^{-1} terms explicitly, and that the order n^{-2} terms could be large if n is not sufficiently large. Finally, this de-biasing removes the principle influence of the prior on the estimators. The prior induces more conservative point estimators since it hedges our bet towards more likely values of f (smaller f). Therefore the bias it induces should only be removed if n is large enough that the bias induced is small. We could only remove the component of the bias associated with the sampling error in W/n if the bias due to the prior is desired.

For all of these reasons the asymptotically debiased estimators should be used with caution. That said, estimating the value of the two components of the bias (bias due to prior, bias due to sampling) is useful as it gives a better understanding of how large the biases in the point estimator are relative to the value of the point estimator, and, as a result,

whether the sample size is sufficiently large.

Another reason that these asymptotic correctors are not essential for point estimation is that we never have more than one sample W (since if we had multiple sampled runs of events they should be collated into one cumulative observation). It follows that we get, at most, one sample from the distribution of possible point estimators. The standard deviation in this distribution is $\mathcal{O}(n^{-1/2})$, so the sampling error is expected to decay slower than the bias as n gets large. Therefore the bias correction is expected to be small relative to the error from finite sample size.

This leads to another set of sample size requirements. In particular, for an edge with true log-odds f it would be natural to require a minimum sample size n such that:

$$n_{\text{min}} = \left\{ n \text{ such that } \frac{2|\sinh(f)|\gamma}{n} < \epsilon|f| \text{ and } \frac{2\sinh(f)}{n} < \epsilon f \right\} \tag{A.82}$$

This leads to the minimum sample size:

$$n_{\text{min}} \geq \frac{\max\{\gamma, 1\}}{\epsilon} \frac{2\sinh(f)}{f}. \tag{A.83}$$

Similarly, requiring that the variance in the point estimators is smaller than a fixed threshold gives:

$$n_{\text{min}} \geq \frac{(2\cosh(f/2))^2}{\epsilon^2}. \tag{A.84}$$

Or, requiring that the signal to noise ratio in the point estimators is greater than a desired SNR gives:

$$n_{\text{min}} \geq \text{SNR} \frac{\sqrt{2}\cosh(f/2)}{|f|} \tag{A.85}$$

Notice that the first two sample size requirements require more samples as f gets larger, while the last sample size requirement requires more samples if f gets either sufficiently

large, or sufficiently small. We do not expect the last requirement to be satisfied by all edges, however we do expect that it is satisfied on at least some edges.

A.9 Point Estimators for the HHD

The HHD is a decomposition of the edge flow. This decomposition is linear. Define the gradient operator G. Then define the transitive projector P_t which is the orthogonal projector onto the range of G. Let $P_c = I - P_t$ be the projector onto the cyclic subspace (subspace perpendicular to the range of G). Then the interesting components of the HHD are:

$$\begin{aligned} r &= (G^T G)^\dagger G^T f \\ f_t &= P_t f \\ f_c &= P_c f \end{aligned} \tag{A.86}$$

where r is the rating, $\dagger$ denotes the pseudo-inverse, f_t is the transitive component, and f_c is the cyclic component.

A.9.1 Point Estimation of Components

Point estimation of the components of the HHD is easy since all of the components are linear functions of the edge flow. Therefore the conditional expectation, and MAP for each components can be computed by applying the HHD to the corresponding estimators for the flow. The variance in the posterior for each component can also be computed directly from the variance in the posterior for the flow. The edges were assumed to be independent, thus $\mathbb{V}[F|n, w, \gamma]$ is diagonal with diagonal entries equal to $\psi^{(1)}(w_k + \gamma) + \psi^{(1)}(n_k - w_k + \gamma)$. Then:

$$\mathbb{V}[AF|n, w, \gamma] = A\mathbb{V}[F|n, w, \gamma]A^T. \tag{A.87}$$

Therefore the MAP estimator and conditional expectation of the components of the HHD are:

$$\begin{aligned}
r_{\text{MAP}}(n,w,\gamma) &= (G^TG)^\dagger G^T f_{\text{MAP}}(n,w,\gamma) \\
f_{t\text{MAP}}(n,w,\gamma) &= P_t f_{\text{MAP}}(n,w,\gamma) \\
f_{c\text{MAP}}(n,w,\gamma) &= P_c f_{\text{MAP}}(n,w,\gamma) \\
r_{\text{exp}}(n,w,\gamma) &= (G^TG)^\dagger G^T f_{\text{exp}}(n,w,\gamma) \\
f_{t\text{exp}}(n,w,\gamma) &= P_t f_{\text{exp}}(n,w,\gamma) \\
f_{c\text{exp}}(n,w,\gamma) &= P_c f_{\text{exp}}(n,w,\gamma)
\end{aligned} \tag{A.88}$$

and the variance in the posterior for each component is:

$$\begin{aligned}
\mathbb{V}[R|n,w,\gamma] &= (G^TG)^\dagger G^T \mathbb{V}[F|n,w,\gamma] G (G^TG)^\dagger \\
\mathbb{V}[F_t|n,w,\gamma] &= P_t \mathbb{V}[F|n,w,\gamma] P_t^T \\
\mathbb{V}[F_c|n,w,\gamma] &= P_c \mathbb{V}[F|n,w,\gamma] P_c^T
\end{aligned} \tag{A.89}$$

A.9.2 Point Estimation of Measures

What remains is to work out point estimators for the measures. The measures are:

$$\begin{aligned}
m_{\text{total}}(f) &= ||f||_2 \\
m_{\text{trans}}(f) &= ||f_t||_2 \\
m_{\text{cyc}}(f) &= ||f_c||_2 \\
m_{\text{relative}}(f) &= ||f_c||_2 / ||f||_2
\end{aligned} \tag{A.90}$$

The measures are not linear functions of the flow, so the MAP estimator, or conditional expectation of the measures is not given by applying the measures to the corresponding estimator for the flow. Nevertheless, the measures may be estimated by applying the

measures to the point estimators for the flow. This is natural as these point estimators are also the point estimators for the components of the HHD measured by the measures. We will show that working with the marginal posterior distributions for the measures can be misleading as uncertainty in the edge flow introduces biases in the marginal posterior distributions of the measures.

Applying the measures to our point estimators for the edge flow is trivial, and consists only of plugging the point estimators in for f in Equation (A.90).

An alternative approach is to approximate the posterior distribution for each measure by sampling edge flows from the posterior distribution of edge flows. This will be discussed in the next section. Asymptotic biases will also be discussed in the next section.

A.10 Sampling Methods

A.10.1 Sampling Edge Flows

Edge flows, F, can be sampled from the posterior distribution of edge flows by sampling win probabilities, P, from the posterior distribution of win probabilities and then computing $F = \text{logit}(P)$. The win probabilities are easy to sample as each win probability is independent of the others and is beta distributed. As a consequence the win probabilities can be sampled from a beta random number generator. To sample the edge flow on edge k from the posterior:

1. Sample $P_k \sim \text{Beta}(w_k + \gamma, n_k - w_k + \gamma)$,
2. Set $F_k = \text{logit}(P_k)$.

A.10.2 Sampling Ratings and Components

The components of the HHD and the ratings are all given by the product of the edge flow with a matrix. This product gives the solution to a linear system whose right hand side depends on the edge flow. Therefore, when the network is sufficiently small the components of the HHD and ratings can be sampled simply by multiplying a sampled F by the appropriate matrix. If the network is too large for direct multiplication to solve the linear systems a linear system can be solved for each sampled edge flow using a linear system solver.

Let $F = [F(1), F(2), ...F(N)]$ where N is a number of realizations and $F(j)$ is a single realization. Then sampled HHD components are given by:

$$\begin{aligned} R &= (G^T G)^\dagger G F \\ F_t &= P_t F \\ F_c &= P_c F = F - F_t. \end{aligned} \tag{A.91}$$

Each sampled rating generates a sampled ranking (list of competitors in order of decreasing rating). These sampled ratings can be compared using either the Kendall tau measure, or Spearman rank correlation. These measure how much variation there is in the ranking across the samples.

Another useful technique is to record the fraction of samples in which a given competitor is assigned a given rank. This gives the posterior distribution for the rank of each competitor. These can be represented conveniently with a matrix whose columns correspond to competitors, and whose rows correspond to rank. Then the entries are the fraction of samples in which competitor i was assigned rank j. If the competitors are ordered in the ranking associated with one of the point estimators for the ratings then

this matrix has entries close to one near its diagonal, and entires close to zero off its diagonal when the ranking is unambiguous. When there is uncertainty in ranking a group of competitors these form a block in the matrix.

A.10.3 Sampling Measures

Once the components of the HHD have been sampled the measures can be sampled by evaluating the measures on each sampled component. This gives a collection of samples of possible values of measures when evaluated on sampled edge flows from the posterior for the edge flow. A histogram of the samples can then give an approximation to the posterior distribution for each measure. Further, the conditional expectation of the value of the measure can be approximated by averaging the sampled values of the measure. Confidence intervals on the value of the measure can be also be established from the samples.

The use of this technique for estimation of the value of the measures is discouraged for the following reason: the posterior distribution for the measures is biased by uncertainty in the posterior distribution for the edge flow. Moreover this bias is not equal amongst the different measures, and typically leads to overestimation of the cyclic component relative to the transitive component.

To see why uncertainty biases point estimates for the measures it is helpful to take a step back and consider the big picture. The absolute measures are distances from either the origin, the cyclic subspace, or the transitive subspace. Let P_S represent an orthogonal projector onto a subspace S. Then let F be an edge flow sampled from some distribution of edge flows (not necessarily the posterior). Then:

$$\mathbb{E}[||P_S F||^2] = ||P_S \mathbb{E}[F]||^2 + \langle P_S, \mathbb{V}[F] \rangle. \tag{A.92}$$

Therefore variance in F biases the expected value of the squared measure associated with the projection onto subspace S. The size of the bias depends on the size of the covariance, and the dimension of the subspace. The larger the dimension of the subspace, the larger $\langle P_S, \mathbb{V}[F] \rangle$ usually is.

Now let's compare two different approaches for estimating the measures. The first is to apply the measure directly to one of the point estimators for the edge flow. The second is to average the measure over the posterior distribution of edge flows. This can be approximated empirically by sampling F, evaluating the measure for each sample, then averaging the sampled measures values.

In the former case the edge flow measured is $f_*(n, W, \gamma)$ where $*$ stand for MAP, exp or either of the asymptotically corrected measures. In each case $\mathbb{E}[f_*(n, W, \gamma)]$ differs from the true f by a term $\mathcal{O}(n^{-1})$ or $\mathcal{O}(n^{-2})$. If an asymptotically corrected estimator is used this bias is $\mathcal{O}(n^{-2})$. Let $b_*(n, \gamma | f)$ be the corresponding bias to $\mathcal{O}(n^{-1})$. The variance in all of the point estimators given f is the same to order n^{-1} and is:

$$\mathbb{V}[f_*(n, W, \gamma)|f] = (2\cosh(f/2))^2 \frac{1}{n} + \mathcal{O}(n^{-2}). \tag{A.93}$$

Thus, using one of the point estimators for the edge flow, the expected error is:

$$\begin{aligned}
\mathbb{E}\left[||P_S f(n,W,\gamma)||^2|f\right] - ||P_S f||^2 &= \left(||P_S(f+b_*(n,\gamma|f))||^2 - ||P_S f||^2\right) \\
&\quad + \langle P_s, \text{diag}((2\cosh(f/2))^2 \frac{1}{n})\rangle + \mathcal{O}(n^{-2}) \\
&= \left(2b_*(n,\gamma|f)^T P_s f + ||P_S b_*(n,W,\gamma|f)||^2\right) \\
&\quad + \sum_{k=1}^{E} \frac{(2\cosh(f_k/2))^2}{n_k} + \mathcal{O}(n^{-2}) \\
&= 2b_*(n,\gamma|f)^T P_s f \\
&\quad + \sum_{k=1}^{E} (P_s)_{kk} \frac{(2\cosh(f_k/2))^2}{n_k} + \mathcal{O}(n^{-2}) \\
&= 2b_*(n,\gamma|f)^T P_s f \\
&\quad + |S| \sum_{k=1}^{E} \frac{(P_s)_{kk}}{\text{trace}(P_S)} \frac{(2\cosh(f_k/2))^2}{n_k} + \mathcal{O}(n^{-2}).
\end{aligned} \tag{A.94}$$

The first term in this equation is the expected error due to the bias in the estimator for the edge flow. The second term is the expected error due to sampling error in the value of the estimator. Both of these errors are $\mathcal{O}(n^{-1})$. For most choices of γ the bias $b_*(n,W,\gamma|f)$ is negative when f is positive, and positive when f is negative, thus the first error is usually negative. The latter error is strictly positive since the diagonal entries of an orthogonal projector are strictly nonnegative, and is equivalent to the dimension of the subspace S times a weighted average of the variance in the posterior for the edge flow on each edge.

Now suppose that, instead of evaluating the measure on a point estimator for the edge flow we average the measure (squared) over its posterior. This is equivalent to finding the average value of the measure applied to F when F is sampled from its posterior. That is $\mathbb{E}[||P_S F||^2|f]$. Note that the posterior for f depends onthe sampled win record, W, which is a random variable.

First, the expected value of F given f is the expected value of F conditioned on sampling $W = w$, averaged over the probability of sampling $W = w$. That is, the expected value of $f_{\text{exp}}(n, W, \gamma)$. Therefore:

$$\mathbb{E}[F|f] = \mathbb{E}[f_{\text{exp}}(n, W, \gamma)|f] = f + b_{\text{exp}}(n, \gamma|f) + \mathcal{O}(n^{-2}). \tag{A.95}$$

Therefore the expected value of F given f has the same bias as the conditional expectation when averaged over possible samples, so is not any more accurate in expectation than the point estimators, and is less accurate (asymptotically) than the corrected estimators.

To compute the variance we use the law of total variance:

$$\mathbb{V}[F|f] = \mathbb{E}[\mathbb{V}[F|W]] + \mathbb{V}[\mathbb{E}[F|W]]. \tag{A.96}$$

The second term is the sampling variance in any of the point estimators:

$$\mathbb{V}[\mathbb{E}[F|W]] = \mathbb{V}[f_{\text{exp}}(n, W, \gamma)|f] = \mathbb{V}[f_*(n, W, \gamma)|f] = (2\cosh(f/2))^2 \frac{1}{n} + \mathcal{O}(n^{-2}). \tag{A.97}$$

The first term is:

$$\begin{aligned}
\mathbb{E}[\mathbb{V}[F|W]] &= \mathbb{E}[\psi^{(1)}(W + \gamma) + \psi^{(1)}(n - W + \gamma)|f] \\
&= \psi^{(1)}(pn + \gamma) + \psi^{(1)}((1 - p)n + \gamma) + \mathcal{O}(n^{-2}) \\
&= \frac{1}{pn + \gamma} + \frac{1}{(1 - p)n + \gamma} + \mathcal{O}(n^{-2}) \\
&= \left(\frac{1}{p} + \frac{1}{1 - p}\right) \frac{1}{n} + \mathcal{O}(n^{-2}) \\
&= \frac{1}{p(1 - p)} \frac{1}{n} + \mathcal{O}(n^{-2}) \\
&= (2\cosh(f/2))^2 \frac{1}{n} + \mathcal{O}(n^{-2})
\end{aligned} \tag{A.98}$$

where the first simplification follows from our general technique for approximating expectations using an asymptotic expansion in the moments. Note that this is the same asymptotic form as the variance in the point estimators. Therefore:

The expected error in the estimated value of the measure $||P_S F||^2$ has the same error due to bias if F is one of the point estimators, or if the measure is averaged over the posterior, however the error due to sampling variance is twice as large if the measure is averaged over the posterior than if it is evaluated at a point estimator since the variance in F when F is sampled from the posterior given f is twice the variance in the point estimators for f given f (to order n^{-2}).

Therefore applying the measures directly to the point estimators will lead to a smaller error due to sampling variance.

One might hope that the shift in the posterior distribution for each measure due to the uncertainty in the edge flow is accompanied by an equivalent increase in the variance in the posterior distribution such that the uncertainty in the posterior for the measure reflects the shift in the measure due to uncertainty. This is emphatically not the case when the uncertainty in F is sufficiently large. In practice we observe that the shift due to the variance is larger than the standard deviation in the posterior for the measures.

A.11 Summary:

A.11.1 Methods:

So far we have developed a series of point estimators for the edge flow, HHD components, and measures. These are complemented by sampling methods which can be used to approximate averages over the posterior distribution, or the posterior distribution itself, of the

rankings and measures. These are summarized below:

Summary: (Point Estimators and Sampling)

Let n_k be the number of events observed between a pair of competitors $i(k), j(k)$. Let w_k be the number of times $i(k)$ beat $j(k)$. Let n be the vector of event counts, and w the vector of win counts. Let γ be the prior parameter (see A.2.1 for instructions on choice of γ). Then:

1. The win probabilities are beta distributed given the data with: $P_k \sim \text{Beta}(w_k + \gamma, n_k - w_k + \gamma)$. Thus the conditional expectation for the win probabilities is: $\mathbb{E}[P_k | n_k, w_k, \gamma] = (w_k + \gamma)/(n_k + 2\gamma)$.

2. The log-odds on edge k are distributed according to:

$$\pi_{F_k}(f|n, w\gamma) = \frac{\text{logistic}(f)^{w_k+\gamma}\text{logistic}(-f)^{n_k-w_k+\gamma}}{B(w_k + \gamma, n_k - w_k + \gamma)}. \tag{A.99}$$

3. Thus, the log-odds can be estimated by:

$$\begin{aligned} f_{\text{MAP}}(n_k, w_k, \gamma) &= \text{logit}(\mathbb{E}[P_k | n_k, w_k, \gamma]) = \ln\left(\frac{w_k + \gamma}{n_k - w_k + \gamma}\right) \\ f_{\text{exp}}(n_k, w_k, \gamma) &= \mathbb{E}[F_k | n_k, w_k, \gamma] = \psi(w_k + \gamma) - \psi(n_k - w_k + \gamma). \end{aligned} \tag{A.100}$$

The conditional expectation is equivalent to:

$$f_{\text{exp}}(n_k, w_k, \gamma) = \sum_{w=0}^{w_k} \frac{1}{w + \gamma} - \sum_{l=0}^{n_k - w_k} \frac{1}{l + \gamma} \tag{A.101}$$

thus can be updated recursively by adding $1/(w_k + \gamma)$ when a win is observed, and subtracting $1/(n_k - w_k + \gamma)$ when a loss is observed. Finally these two estimators

converge to each other in the limit of large w and $n - w$ with:

$$\begin{aligned} f_{\text{exp}}(n, w, \gamma) - f_{\text{MAP}}(n, w, \gamma) &= \frac{1}{2} \frac{2w - n}{(w + \gamma)(n - w + \gamma)} \\ &\quad + \mathcal{O}((w + \gamma)^{-2}) + \mathcal{O}((n - w + \gamma)^{-2}) \\ f_{\text{exp}}(n, w, \gamma) - f_{\text{MAP}}(n, w, \gamma) &\in \frac{1}{2} \frac{2w - n}{(w + \gamma)(n - w + \gamma)} \\ &\quad + \left[\frac{-1}{2(w + \gamma)}, \frac{1}{2(n - w + \gamma)} \right]. \end{aligned} \tag{A.102}$$

4. The posterior distribution for the log-odds is unimodal, has convex negative log-likelihood, and has tails that decay exponentially according to:

$$\begin{aligned} \lim_{f \to \infty} \pi_F(f|n, w, \gamma) &\propto \lim_{f \to \infty} \exp(-(n - w + \gamma)f) \\ \lim_{f \to -\infty} \pi_F(f|n, w, \gamma) &\propto \lim_{f \to -\infty} \exp(-(w + \gamma)f) \end{aligned} \tag{A.103}$$

thus skews negative if fewer wins than losses are observed, and skews positive if fewer losses than wins are observed. The variance in the posterior distribution for the log-odds is:

$$\begin{aligned} \mathbb{V}[F|n, w, \gamma] &= \psi^{(1)}(w + \gamma) + \psi^{(1)}(n - w + \gamma) \\ &\sim \frac{1}{w + \gamma} + \frac{1}{n - w + \gamma} + \mathcal{O}((w + \gamma)^{-2}) + \mathcal{O}((n - w + \gamma)^{-2}). \end{aligned} \tag{A.104}$$

Therefore the variance in the posterior is only small if both the observed number of wins and the observed number of losses is large. Sample size requirements can be derived by bounding this variance, or by bounding the expected size of this variance relative to the expected size of the point estimators when $P = \text{logistic}(F)$ is sampled from the prior $\text{Beta}(\gamma, \gamma)$.

5. In the limit of large sample size the expected error in the point estimators given the true f is:

$$\begin{aligned}\mathbb{E}[f_{\text{MAP}}(n,W,\gamma)|f] - f &= -2\sinh(f)\frac{(\gamma-1/2)}{n} + \mathcal{O}(n^{-2}) \\ \mathbb{E}[f_{\text{exp}}(n,W,\gamma)|f] - f &= -2\sinh(f)\frac{(\gamma-1)}{n} + \mathcal{O}(n^{-2}).\end{aligned} \tag{A.105}$$

and the variance in the point estimators is:

$$\begin{aligned}\mathbb{V}[f_{\text{MAP}}(n,W,\gamma)|f] &= (2\cosh(f/2))^2\frac{1}{n} + \mathcal{O}(n^{-2}) \\ \mathbb{V}[f_{\text{exp}}(n,W,\gamma)|f] &= (2\cosh(f/2))^2\frac{1}{n} + \mathcal{O}(n^{-2}).\end{aligned} \tag{A.106}$$

Therefore the asymptotic bias in the point estimators is $\mathcal{O}(n^{-1})$, while the standard deviation in the point estimators is $\mathcal{O}(n^{-1/2})$. Therefore, in the limit of large sample size the bias in the expected value of the point estimators vanishes faster than the standard deviation in the point estimators. This bias can be corrected with the point estimators:

$$\begin{aligned}f_{\text{MAP}*}(n,W,\gamma) &= f_{\text{MAP}}(n,W,\gamma) + 2\sinh(f_{\text{MAP}}(n,W,\gamma))\frac{(\gamma-1/2)}{n} \\ f_{\text{exp}*}(n,W,\gamma) &= f_{\text{exp}}(n,W,\gamma) + 2\sinh(f_{\text{exp}}(n,W,\gamma))\frac{(\gamma-1)}{n}\end{aligned}. \tag{A.107}$$

for which the expected error is order n^{-2} and the variance is unchanged to order n^{-1}. Additional sampling size requirements can be introduced to ensure that the expected bias and uncertainty are sufficiently small.

6. The MAP estimators and conditional expectations of the components of the HHD are

given by:

$$\begin{aligned} r_{\text{MAP}}(n, w, \gamma) &= (G^T G)^\dagger G^T f_{\text{MAP}}(n, w, \gamma) \\ f_{t\text{MAP}}(n, w, \gamma) &= P_t f_{\text{MAP}}(n, w, \gamma) \\ f_{c\text{MAP}}(n, w, \gamma) &= P_c f_{\text{MAP}}(n, w, \gamma) \\ r_{\text{exp}}(n, w, \gamma) &= (G^T G)^\dagger G^T f_{\text{exp}}(n, w, \gamma) \\ f_{t\text{exp}}(n, w, \gamma) &= P_t f_{\text{exp}}(n, w, \gamma) \\ f_{c\text{exp}}(n, w, \gamma) &= P_c f_{\text{exp}}(n, w, \gamma) \end{aligned} \tag{A.108}$$

and the variance in the posterior for each component is:

$$\begin{aligned} \mathbb{V}[R|n, w, \gamma] &= (G^T G)^\dagger G^T \mathbb{V}[F|n, w, \gamma] G (G^T G)^\dagger \\ \mathbb{V}[F_t|n, w, \gamma] &= P_t \mathbb{V}[F|n, w, \gamma] P_t^T \\ \mathbb{V}[F_c|n, w, \gamma] &= P_c \mathbb{V}[F|n, w, \gamma] P_c^T \end{aligned} \tag{A.109}$$

7. The measures can be estimated by evaluating:

$$\begin{aligned} m_{\text{total}}(f) &= ||f||_2 \\ m_{\text{trans}}(f) &= ||f_t||_2 \\ m_{\text{cyc}}(f) &= ||f_c||_2 \\ m_{\text{relative}}(f) &= ||f_c||_2 / ||f||_2 \end{aligned} \tag{A.110}$$

where either f_{MAP} or f_{exp} is used for the edge flow f. This maintains consistency with the estimators for the components of the HHD. Alternatively the measures can be estimated by sampling from the posterior distribution of edge flows (see below), evaluating the measure for each sampled edge flow, and averaging. This is discouraged as it is expected to double the bias in the estimated measure due

to sampling uncertainty. The variance in the posterior can also be estimated by sampling.

8. We can sample from the posterior for the edge flow by sampling $P_k \sim \text{Beta}(w_k + \gamma, n_k - w_k + \gamma)$, then letting $F_k = \text{logit}(P_k)$. By multiplying by the appropriate matrix this gives samples from the posterior distribution for each component of the HHD. By evaluating the measures at the sampled F this gives samples from the posterior distribution for the measures. These could be compared to samples from the prior (draw P_k from $\text{Beta}(\gamma, \gamma)$) to understand how the data has changed the distribution of each measure. The Kolmogorov distance, or an estimator for the KL distance, can be used to quantify how much has been learned about the measures from the data.

9. For each sampled F a rating R and associated ranking can be computed. This generates sampled ratings from the posterior distribution for the ratings. The Spearman rank correlation and Kendall tau measure can be computed for the list of sampled ratings in order to quantify our certainty in the ratings. Histograms for the ratings of each competitor can be used to approximate the posterior distribution of ratings, and can be collected into a confusion matrix.

A.11.2 Limitations and Challenges:

The summary presented above gives a complete estimation protocol for estimating the edge flow, components of the HHD, and measures. The estimation protocol has two primary limitations:

1. The tails of the posterior for F decay exponentially with rates proportional to the number of observed wins for negative F, and number observed losses for positive F. Therefore, if the observed number of wins, or observed number of losses, is

small then the posterior has a large variance, and at least one slowly decaying tail. That is: if we do not observe many losses it is hard to put an upper bound on the log-odds of winning, and if we do not observe many wins it is hard to put a lower bound on the log-odds of losing. Thus rare events make point estimation difficult. This difficulty can lead to surprising limitations. For example, suppose we observe a competitor win 10 of 10 events against an opponent. Then it is clear that they likely have a win probability greater than $1/2$. But how much greater? This question is not easy to answer since we have not observed any losses, so have no data to bound the win probability above. Instead we are forced to rely entirely on the prior to put an upper bound on the estimated win probability. This is not a serious problem when estimating the win probabilities since probabilities are bounded above and below by 1 and 0. It is a problem when estimating the log-odds since the log-odds are unbounded (map 1 tp infinity and 0 to negative infinity). Therefore we should not expect to be able to make confident point predictions of the value of the log-odds without sufficiently many wins and losses. As a result, even if we see A beat B 10 out of 10 games, B beat C 10 out of 10 games, and C beat A 10 out of 10 games, we cannot reach a confident estimate of how cyclic the system is. It is clearly cyclic, but how cyclic?

2. The analysis described above is entirely Bayesian. The point estimators are limited in that they are points representing a distribution, but even if we sample from, or compute the posterior explicitly, the posterior is only appropriate for answering some questions. The posterior gives the probability that a certain edge flow is the true edge flow (under the chosen prior - which is itself a modelling assumption) given the data. When the data is insufficient to keep the variance in the posterior small a wide range of possible edge flows could correspond to the data, and the shape of the posterior

depends heavily on the choice of prior. The prior assumed the edge flow on each edge is independent of the flow on the other edges, and that logistic(F) is distributed according to Beta(γ, γ). The more the posterior resembles the prior the more our predictions based on the posterior reflect the prior assumptions. We have shown that assuming independent edges promotes cyclic competition as the cyclic subspace usually has larger dimension than the transitive subspace. As a result, even if the true edge flow is on the transitive subspace, but the too few events are observed to resolve the posterior, the point estimators for the measures will usually predict a large cyclic component, since, given limited data, the average edge flow that could correspond to the data has a large cyclic component. This is a fundamental limitation to the Bayesian approach. It can only tell us what edge flows might correspond to the data - thus, since most edge flows have a larger cyclic component, when we are uncertain what edge flow might correspond to the data we find that the most edge flows that could match the data are moderately to highly cyclic.

Moving beyond these limitations requires asking different questions. These questions are:

1. Is there a perfectly transitive edge flow that could plausibly match the data observed? Is there a perfectly cyclic edge flow that could plausibly match the data observed?

2. What is the smallest intransitivity such that there is an edge flow that could plausibly match the data? What is the smallest transitivity such that there is an edge flow that could plausibly match the data?

If we can answer these questions then we can move beyond the two limitations listed above as an edge flow could plausibly correspond to the observed data without resembling

most of the other edge flows that could match the data, and we may be able to bound the measures without needing a precise estimate of the measures. Our approach to answering these questions is described in the next two sections.

Appendix B

Hypothesis Testing Details

In this section we develop tools for testing the hypotheses: H_t,, the tournament is perfectly transitive and H_c the model is perfectly cyclic, against the null hypothesis H_0 the tournament is not necessarily perfectly transivie or perfectly cyclic. In terms of the edge flow these hypotheses are:

1. H_t: $f \in \text{range}\{G\} = \text{null}\{C\}$

2. H_c: $f \in \text{null}\{G^T\} = \text{range}\{C^T\}$.

Notice that both of these hypotheses are subsets of H_0. Therefore H_t and H_c can be considered models that are nested within the generic model in which f can be any vector in $\mathbb{R}^E$. In this context a natural way to test hypothesis is to compute either the likelihood-ratio or the AIC (Aikake Information Criterion). Other test statistics will be considered in subsequently.

The likelihood-ratio is a test statistic, that is, a function of the sampled data that used to test a hypothesis. Let $\mathcal{F}$ be a subset of $\mathbb{R}^E$. Let $\mathcal{L}(f|n, w, \gamma)$ be the likelihood $F = f$ given

$W = w$. Then the likelihood-ratio is defined:

$$\begin{aligned} \mathrm{LR}_{\mathcal{F}}(W|n,\gamma) &= -2\ln\left(\frac{\sup_{f\in\mathcal{F}}\mathcal{L}(f|n,W,\gamma)}{\sup_{f\in\mathbb{R}^E}\mathcal{L}(f|n,W,\gamma)}\right) \\ &= 2\left(\ln(\mathcal{L}(f_{MAP}(n,W,\gamma)|n,W,\gamma)) - \sup_{f\in\mathcal{F}}\ln(\mathcal{L}(f|n,W,\gamma))\right). \end{aligned} \tag{B.1}$$

The likelihood-ratio is the difference between the log-likelihood of the MAP estimator and the MAP estimator constrained to the space $\mathcal{F}$. It is always positive since $\mathcal{F} \subseteq \mathbb{R}^E$. A large likelihood-ratio (large the log-likelihood difference) indicates a large discrepancy between the data and the best model given the hypothesis relative to what is expected under the null hypothesis.

A limitation of the likelihood-ratio is that it does not account for the difference in number of degrees of freedom between H_0 and the hypothesis we are testing. Both H_t and H_c have fewer than E degrees of freedom, thus are expected to provide a worse fit to the data. Let $|\mathcal{F}|$ be the cardinality of $\mathcal{F}$ (number of degrees of freedom in f given hypothesis H). Then the AIC is defined:

$$\mathrm{AIC}_{\mathcal{F}}(W|n,\gamma) = 2\left(|\mathcal{F}| - \ln(\sup_{f\in\mathcal{F}}\mathcal{L}(f|n,W,\gamma))\right). \tag{B.2}$$

Hypothesis can then be compared by comparing AIC values.

B.1 Constrained MAP Estimation

A first step towards computing either the log likelihood ratio or the AIC is to find the MAP estimate for the edge flow constrained to a subspace. This is done numerically.

The function to be maximized is:

$$\ln\left(\pi_F(f|n,w,\gamma)\right) = \sum_{k=1}^{E} \ln(\pi_{F_k}(f|n,w,\gamma)) \tag{B.3}$$

Since $\ln(\pi_{F_k}(f|n,w,\gamma))$ equals $-\sum_{k=1}^{E}(w_k+\gamma)\ln\left(1+\exp(-f_k)\right)+(n_k-w_k+\gamma)\ln\left(1+\exp(f_k)\right)$ up to addition by a constant an equivalent problem is to minimize:

$$\begin{aligned}\ln\left(\pi_F(f|n,w,\gamma)\right) = &\sum_{k=1}^{E}(w_k+\gamma)\ln\left(1+\exp(-f_k)\right) \\ &+ (n_k - w_k + \gamma)\ln\left(1+\exp(f_k)\right).\end{aligned} \tag{B.4}$$

So, given a linear subspace $\mathcal{F}$ the constrained MAP or MLE estimate is given by solving:

$$\text{argmin}_{f\in\mathcal{F}}\left\{\sum_{k=1}^{E}(w_k+\gamma)\ln\left(1+\exp(-f_k)\right) + (n_k - w_k + \gamma)\ln\left(1+\exp(f_k)\right)\right\} \tag{B.5}$$

Since all affine subspaces are convex, and since the cost function is convex (log-posterior or log-likelihood), this is a convex optimization problem. Note that the MLE estimate is given by solving the same type of problem as the MAP estimate, only with $\gamma = 1$.

B.1.1 Comparison to Least Squares Rating

What is the MAP rating if we constrain to the transitive subspace? This is the set of ratings r that solve the optimization problem given in Equation (B.5) with $f_k = r_{i(k)} - r_{j(k)}$. How do these ratings compare to log least-squares ratings?

Log least-squares ratings are given by solving for a rating r that minimizes the (possibly weighted) least-squares distance between Gr and f for some edge flow f. Least squares

ratings can include a regularization term to avoid excessively large ratings.

Finding the MAP rating when f is constrained to the transitive subspace requires minimizing the posterior over the transitive subspace. The posterior is not quadratic, however may be approximated by a quadratic function near $f_{MAP}(n, w, \gamma)$. Therefore, if $f_{MAP}(n, w, \gamma)$ is close to the transitive subspace the value of the log-likelihood can be approximated with a quadratic function. This produces a least squares problem that is an approximation to the true optimization problem.

To approximate the log-posterior with a quadratic function Taylor expand the log-posterior about f_{MAP}. The gradient of the log-posterior is given by Equation (A.25):

$$\partial_k(-\ln(\mathcal{L}(f|n, w, \gamma))) = -(w_k + \gamma) + (n_k + 2\gamma)\text{logistic}(f_k). \tag{B.6}$$

Since the k^{th} partial of the log-likelihood only depends on f_k the Hessian of the cost function is diagonal with diagonal entries:

$$\begin{aligned}
\partial_k^2(-\ln(\mathcal{L}(f|n, w, \gamma))) &= (n_k + 2\gamma)\partial_{f_k}\text{logistic}(f_k) = (n_k + 2\gamma)\frac{\exp(-f_k)}{(1 + \exp(-f_k))^2} \\
&= (n_k + 2\gamma)\frac{1}{1 + \exp(-f_k)}\frac{\exp(-f_k)}{1 + \exp(-f_k)} \\
&= (n_k + 2\gamma)\text{logistic}(f_k)\text{logistic}(-f_k) \\
&= (n_k + 2\gamma)\text{logistic}(f_k)(1 - \text{logistic}(f_k)).
\end{aligned} \tag{B.7}$$

Notice that this is the variance in the binomial distribution for W if the win probability is set to logistic(f_k) and $n_k + 2\gamma$ games are observed. The MAP estimator for f is equivalent to logit$(\mathbb{E}[P|n, w, \gamma])$ therefore logistic$(f_{\text{MAP}}(n, w, \gamma)) = (w + \gamma)/(n + 2\gamma)$. Therefore,

the Hessian evaluated at $f_{\text{MAP}}(n, w, \gamma)$ is:

$$H(n, w, \gamma) = \text{diag}\left(\frac{(w+\gamma)(n-w+\gamma)}{n+2\gamma}\right). \tag{B.8}$$

It follows that the quadratic approximation to the negative log-likelihood about the MAP estimate, $f_{\text{MAP}}(n, w, \gamma)$, is:

$$\begin{aligned} -\ln(\mathcal{L}(f|n, w, \gamma)) \simeq & -\ln(\mathcal{L}(f_{\text{MAP}}(n, w, \gamma)|n, w, \gamma)) \\ & + \frac{1}{2}\sum_{k=1}^{E} \frac{(w_k+\gamma)(n_k-w_k+\gamma)}{n_k+2\gamma}(f - f_{\text{MAP}}(n, w, \gamma))^2. \end{aligned} \tag{B.9}$$

Therefore, the log-least squares approximation to the MAP estimate of the ratings for f constrained to the transitive subspace is:

$$\text{argmin}_{r|\sum_{j=1}^{m} r_j=0} \left\{ \sum_{k=1}^{E} \frac{(w_k+\gamma)(n_k-w_k+\gamma)}{n_k+2\gamma} \left((r_{i(k)} - r_{j(k)}) - \ln\left(\frac{w_k+\gamma}{n_k-w_k+\gamma}\right)\right)^2 \right\} \tag{B.10}$$

or, in terms of $f_{\text{MAP}}(n, w, \gamma)$:

$$\text{argmin}_{r|\sum_{j=1}^{m} r_j=0} \left\{ \sum_{k=1}^{E} \frac{n_k+2\gamma}{\cosh(f_{\text{MAP}}(n_k, w_k, \gamma)/2)^2} \left((r_{i(k)} - r_{j(k)}) - f_{\text{MAP}}(n_k, w_k, \gamma)\right)^2 \right\}. \tag{B.11}$$

Here γ acts to both increase the weights in the sum of squares and to reduce the magnitude of the MAP estimate of f. This in turn reduces the estimated ratings, so acts like a regularizer on the ratings. The least squares approximation to the MAP ratings with f constrained to the transitive subspace can be written more succinctly as the solution to:

$$\text{argmin}_{r|\sum_{j=1}^{m} r_j=0} \left\{ ||\mathbb{V}(W|n, p = \mathbb{E}[P|n, w, \gamma])(Gr - \text{logit}(\mathbb{E}[P|n, w, \gamma]))||^2 \right\}. \tag{B.12}$$

This only differs from the MAP ratings by the weights associated with the variance.

B.1.2 The Likelihood-Ratio as a KL divergence

The log-likelihood of an edge flow f and the MLE edge flow f_{MLE} are:

$$\begin{aligned}
\ln(\mathcal{L}(f|w,n)) &= \sum_{k=1}^{|E|} \ln \binom{n_k}{w_k} - w_k \ln(1+\exp(-f_k)) - (n_k - w_k)\ln(1+\exp(f_k)). \\
\ln(\mathcal{L}(f_{\text{MLE}}|w,n)) &= \sum_{k=1}^{|E|} \ln \binom{n_k}{w_k} + w_k \ln\left(\frac{w_k}{n_k}\right) + (n_k - w_k)\ln\left(\frac{n_k - w_k}{n_k}\right)
\end{aligned} \tag{B.13}$$

Therefore the log likelihood ratio is:

$$\begin{aligned}
&-2\sum_{k=1}^{E} w_k \log\left(\frac{\text{logistic}(f_k)}{w_k/n_k}\right) + (n_k - w_k)\log\left(\frac{\text{logistic}(-f_k)}{(n_k-w_k)/n_k}\right) \\
&= -2\sum_{k=1}^{E} n_k \left[\frac{w_k}{n_k}\log\left(\frac{\text{logistic}(f_k)}{w_k/n_k}\right) + \frac{(n_k-w_k)}{n_k}\log\left(\frac{\text{logistic}(-f_k)}{(n_k-w_k)/n_k}\right)\right] \\
&= 2\sum_{k=1}^{E} n_k D_{\text{KL}}\left(\left\{\frac{w_k}{n_k}, \frac{n_k-w_k}{n_k}\right\} || \{\text{logistic}(f_k), \text{logistic}(-f_k)\}\right)
\end{aligned} \tag{B.14}$$

which is the twice a weighted sum of the KL-divergence between the observed win frequency, and the predicted win frequencies given edge flow f on each edge, where the weights are the number of events observed on each edge.

B.2 Test Statistics

Now, to compute the likelihood-ratio or AIC for either the perfectly transitive or perfectly cyclic hypothesis minimize the negative log-likelihood constrained to the appropriate sub-

space. An initial guess at the constrained MAP edge flow can be given by the corresponding component of the MAP estimate of the edge flow. Then by evaluating the log-likelihood, or by comparing the AIC of the constrained models to the AIC of the MAP estimate, we can sustain or reject the perfectly cyclic or transitive hypotheses.

In addition to the log-likelihood and AIC other test statistics can be used to evaluate whether or not the data could have been plausibly generated by a given edge flow. Since we are already computing the log-likelihood and AIC it is natural to seek test statistics that have the following property:

A natural alternative test statistic is the probability of sampling a win record W given n and f that would be more or equally surprising as the true win record:

$$\text{p-value}(f|n,w) = \Pr\left\{\Pr\{W|n,\text{logistic}(f)\} \leq \Pr\{w|n,\text{logistic}(f)\}\right\}. \tag{B.15}$$

The p-value of a given edge-flow can be easily estimated numerically by sampling. Given an edge flow f set the win probabilities to $\text{logit}(f)$ then sample $W \sim \text{binomial}(n,p)$. Evaluate the probability of sampling w given n,p and evaluate the probability of sampling W for each sample. Then count the fraction of the samples that were less likely than w.

Appendix C

Estimation Details using a Poisson Scoring Model

C.1 The Model

Here we will introduce a simple probabilistic model for baseball. Similar models could be developed for other sports in which the team with the most points wins, and where the game is divided into multiple scoring periods with overtime used to resolve ties. The goal of this model is to allow us to estimate baseball win probabilities from observed line scores (runs per inning).

The model is as follows:

1. Let $R_A(j)$ be the number of runs scored by team A against B in inning j. Similarly let $R_B(j)$ be the number of runs scored by team B against A in inning j.

2. Assume that $R_A(j)$ are all drawn identically and independently from a Poisson distribution with parameter λ_{AB}. Similarly assume that $R_B(j)$ are all drawn identically

and independently from a Poisson distribution with parameter λ_{BA}.

3. Both teams play at least nine innings. If, after nine innings one of the teams has accumulated more wins then they are the winner. If they are still tied after nine innings they continue to play extra innings until one team has a lead, at which point the game ends and the team with the lead is the winner.

C.2 Win Probabilities

Given λ_{AB} and λ_{BA} what is the probability team A beats team B (denoted $A > B$)?

There are two possible ways in which A can beat B. Either A wins in nine innings and the game does not go into extra innings, or A wins in extra innings. Let N be the number of innings played. Since these two outcomes are disjoint so:

$$\text{Pr}\{A > B\} = \text{Pr}\{A > B \cap N = 9\} + \text{Pr}\{A > B \cap N > 9\}. \tag{C.1}$$

In order for the game to go into extra innings the two teams must have been tied at the 9th inning. Therefore the probability A beats B in extra innings is independent of the scores of the two teams at the 9th inning. Moreover, in order for A to beat B in inning $N = n > 9$ the teams must have been tied up to inning $n - 1$, so the probability A beats B in inning n is independent of the teams scores up to inning $n - 1$. That is, all a team needs to do to win in extra-innings is win the last of the extra-innings. Since we assumed the runs scores in each inning are i.i.d this is the same as the probability that A scores more than B

in any single inning conditioned on them not tying. Therefore:

$$\begin{aligned}\Pr\{A > B\} =& \Pr\{\sum_{j=1}^{9}(R_A(j) - R_B(j)) > 0\} \\ &+ \Pr\{R_A(1) > R_B(1) | R_A(1) \neq R_B(1), N > 9\}\Pr\{N > 9\}.\end{aligned} \tag{C.2}$$

Let $S_A(j) = \sum_{i=1}^{j} R_A(j)$ and $S_B(j) = \sum_{i=1}^{j} R_B(j)$ be the cumulative scores of each team after j innings. Let $L_{AB}(j) = S_A(j) - S_B(j)$ be team A's lead at inning j. Now:

$$\Pr\{A > B\} = \Pr\{L_{AB}(9) > 0\} + \Pr\{L_{AB}(1) > 0 | L_{AB}(1) \neq 0\}\Pr\{L_{AB}(9) = 0\}. \tag{C.3}$$

So, in order to compute the win probability all we need is the cumulative distribution function of $L_{AB}(j)$ at $j = 1$ and $j = 9$.

First, $L_{AB}(1) = R_A(1) - R_B(1)$. Both $R_A(1)$ and $R_B(2)$ are Poisson, and they are independent of each other. Therefore $L_{AB}(1)$ is Skellam distributed:

$$\Pr\{L_{AB}(1) = l\} = e^{-(\lambda_{AB}+\lambda_{BA})}\left(\frac{\lambda_{AB}}{\lambda_{BA}}\right)^{l/2} I_l(2\sqrt{\lambda_{AB}\lambda_{BA}}). \tag{C.4}$$

Here $I_\nu(x)$ is the modified Bessel function of the first kind. Unfortunately the cdf of the Skellam distribution is not known in closed form. Therefore, in its simplest form:

$$\Pr\{L_{AB}(1) > 0\} = e^{-(\lambda_{AB}+\lambda_{BA})}\sum_{l=1}^{\infty}\left(\frac{\lambda_{AB}}{\lambda_{BA}}\right)^{l/2} I_l(2\sqrt{\lambda_{AB}\lambda_{BA}}). \tag{C.5}$$

The probability that the two teams tie in a single inning is:

$$\Pr\{L_{AB}(1) = 0\} = e^{-(\lambda_{AB}+\lambda_{BA})} I_0(2\sqrt{\lambda_{AB}\lambda_{BA}}). \tag{C.6}$$

Therefore the probability they don't tie is:

$$\Pr\{L_{AB}(1) \neq 0\} = 1 - e^{-(\lambda_{AB}+\lambda_{BA})} I_0(2\sqrt{\lambda_{AB}\lambda_{BA}}). \tag{C.7}$$

This gives the probability that team A scores more than team B in any inning, thus the probability of A winning in extra innings if the game goes to extra innings. To find out the probability the game goes into extra innings, or that A wins in 9 innings, we need to know how $L_{AB}(9)$ is distributed.

The sum of independent Poisson distributed random variables is Poisson distributed with mean equal to the sum of the means of each random variable. Therefore $S_A(9)$ is Poisson distributed with mean $9\lambda_{AB}$ and $S_B(9)$ with mean $9\lambda_{BA}$. Therefore $L_{AB}(9)$ is also Skellam distributed, only with means $9\lambda_{AB}$ and $S_B(9)$. It follows that:

$$\Pr\{L_{AB}(9) = 0\} = e^{-9(\lambda_{AB}+\lambda_{BA})} I_0(18\sqrt{\lambda_{AB}\lambda_{BA}}) \tag{C.8}$$

and:

$$\Pr\{L_{AB}(9) > 0\} = \sum_{l=1}^{\infty} e^{-9(\lambda_{AB}+\lambda_{BA})} \left(\frac{\lambda_{AB}}{\lambda_{BA}}\right)^{l/2} I_l(18\sqrt{\lambda_{AB}\lambda_{BA}}). \tag{C.9}$$

So, letting:

$$\text{Skellam}(l|\mu,\lambda) = e^{-(\lambda+\mu)} \left(\frac{\lambda}{\mu}\right)^{l/2} I_l(2\sqrt{\lambda\mu}). \tag{C.10}$$

Then, the probability A beats B is:

$$\begin{aligned}\Pr\{A > B\} = &\sum_{l=1}^{\infty} \text{Skellam}\left(l|9\lambda_{AB}, 9\lambda_{BA}\right) \\ &+ \frac{\text{Skellam}\left(0|9\lambda_{AB}, 9\lambda_{BA}\right)}{1 - \text{Skellam}\left(0|\lambda_{AB}, \lambda_{BA}\right)} \sum_{l=1}^{\infty} \text{Skellam}\left(l|\lambda_{AB}, \lambda_{BA}\right)\end{aligned} \tag{C.11}$$

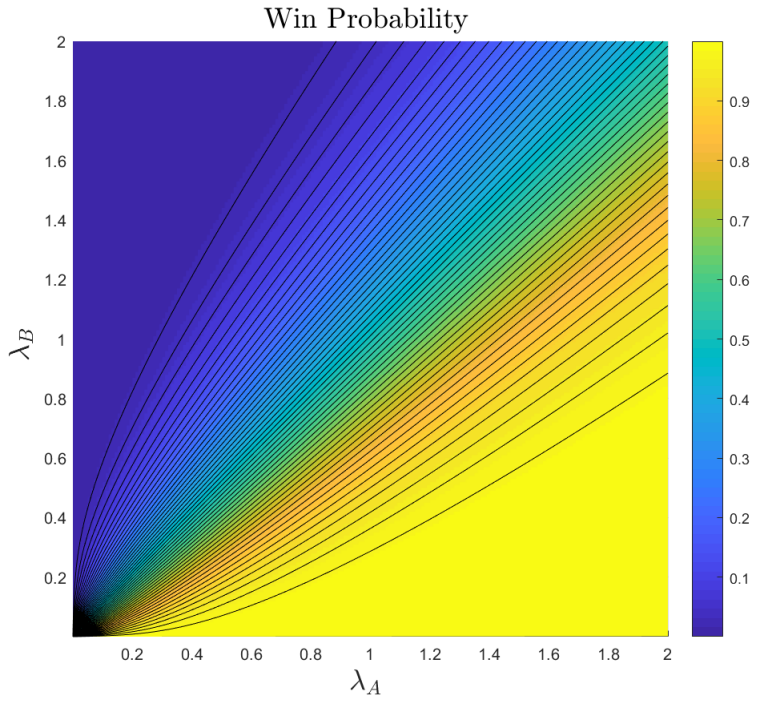

Figure C.1: Win probability for team A against team B if team A scores with rate λ_A against team B, and team B scores with rate λ_B. Blue represents low probability, yellow represents high probability. Notice that the win probability function is a smooth step function, with low probability if λ_B is sufficiently greater than λ_A, and high probability otherwise.

The win probability function defined by Equation (C.11) is illustrated in Figure C.1.

C.3 Estimation

C.3.1 Estimating λ

How can we estimate λ_{AB} and λ_{BA} given data?

Line score data (runs scored in each inning for each team) is widely available for MLB games, and has been collated on sabermetrics sites. For example, retrosheet.org provides the line scores of every MLB game in 2019, as well as extensive historical data.

In order to estimate λ_{AB} and λ_{BA} we assume that these are fixed over the course of a season. That is, we assume that the expected number of runs hit by team A's offense against team B's defense is constant over the season.

Let $n_{AB}(y)$ be the total number of innings seen played between A and B in year y. Let $\{r_{AB}(y)_j\}j = 1^{n_{AB}(y)}$ be the runs scored by A against B in the j^{th} inning they played during year y. Since it is assumed that each inning is independent of each other inning, and the runs scored in each inning are Poisson:

$$\Pr\{R_{AB}(y) = r_{AB}(y)\} = \prod_{j=1}^{n_{AB}(y)} \frac{\lambda_{AB}(y)^{r_{AB}(y)_j}}{r_{AB}(y)_j!} e^{-\lambda_{AB}(y)}. \tag{C.12}$$

This is the likelihood.

We now need a prior on $\lambda_{AB}(y)$. We will assume that all $\lambda_{ij}(y)$ are sampled i.i.d from a gamma distribution with parameters α and β since the gamma distribution is the conjugate prior to the Poisson distribution. Then:

$$\Pr\{\Lambda_{AB}(y) = \lambda\} = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}. \tag{C.13}$$

Then the posterior distribution for $\Lambda_{A,B}(y)$ is proportional to:

$$\Pr\{\Lambda_{AB}(y) = \lambda | r_{AB}(y), \alpha, \beta\} \propto \frac{\beta^\alpha}{\Gamma(\alpha)} \left[\prod_{j=1}^{n_{AB}(y)} \lambda^{r_{AB}(y)_j} e^{-\lambda} \right] \lambda^{\alpha-1} e^{-\beta\lambda} \tag{C.14}$$

or, equivalently:

$$\Pr\{\Lambda_{AB}(y) = \lambda | r_{AB}(y), \alpha, \beta\} \propto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\sum_{j=1}^{n_{AB}(y)} r_{AB}(y)_j + \alpha - 1} e^{-(n_{AB}(y)+\beta)\lambda}. \tag{C.15}$$

Therefore the posterior distribution for Λ is a gamma distribution with parameters

$\sum_{j=1}^{n_{AB}(y)} r_{AB}(y)_j + \alpha$ and $n_{AB}(y) + \beta$:

$$\Lambda_{AB}(y) \sim \text{Gamma}\left(\sum_{j=1}^{n_{AB}(y)} r_{AB}(y)_j + \alpha, \quad n_{AB}(y) + \beta\right). \tag{C.16}$$

This gives an intuitive interpretation of the prior parameters. The first prior parameter, α, is a fictitious number of runs added to the total number of runs scored by A against B in year y. The second parameter, β, is a fictitious number of innings across which the fictitious runs occurred. Moreover, since the posterior is a gamma distribution it is easy to sample from.

The mean, variance, and mode of gamma distributions are easy to evaluate. The mode and mean give the MAP estimator for $\Lambda_{AB}(y)$ and its conditional expectation. The variance is the uncertainty in these estimates:

$$\begin{aligned}
\lambda_{MAP}(r, n, \alpha, \beta) &= \frac{\sum_{j=1}^{n} r(j) + \alpha - 1}{n + \beta} \\
\lambda_{exp}(r, n, \alpha, \beta) = \mathbb{E}[\Lambda | r, n, \alpha\beta] &= \frac{\sum_{j=1}^{n} r(j) + \alpha}{n + \beta} = \lambda_{MAP}(r, n, \alpha, \beta) + \frac{1}{n + \beta} \\
\text{Var}[\Lambda | r, n, \alpha, \beta] &= \frac{\sum_{j=1}^{n} r(j) + \alpha}{(n + \beta)^2} = \frac{\mathbb{E}[\Lambda | r, n, \alpha\beta]}{n + \beta}.
\end{aligned} \tag{C.17}$$

C.3.2 Estimating the prior parameters

How should we estimate the prior parameters α, β?

Assume that the prior parameters are fixed for some series of seasons Y. This could be a single season or multiple seasons. Then the probability of observing the line scores of

each pair of teams during those seasons is:

$$\Pr\{\text{data}(Y)|\alpha,\beta\} = \prod_{y\in Y}\prod_{(ij)\in VE(y)}\int_0^\infty \left[\prod_{k=1}^{n_{ij}(y)} \Pr\{R_{ij}(y)_k = r_{ij}(y)_k|\lambda_{ij}(y)=\lambda\}\right] \times \Pr\{\lambda_{ij}(y)=\lambda|\alpha,\beta\}d\lambda. \tag{C.18}$$

Here $VE(Y)$ denotes the set of all pairs of teams who competed in year y (this is the set of edges of the network associated with year y).

The inner pair of probabilities are the likelihood we already evaluated so the integral is:

$$\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{1}{\prod_{k=1}^n r_k!}\int_0^\infty \lambda^{\sum_{k=1}^n r_k+\alpha-1}e^{-(n+\beta)\lambda}d\lambda = \frac{\Gamma\left(\sum_{k=1}^n r_k+\alpha\right)}{\prod_{k=1}^n r_k!\,\Gamma(\alpha)}\beta^\alpha(n+\beta)^{-\left(\sum_{k=1}^n r_k+\alpha\right)}. \tag{C.19}$$

Therefore:

$$\Pr\{\text{data}(Y)|\alpha,\beta\} = \prod_{y\in Y,(ij)\in VE(Y)}\frac{\Gamma\left(\sum_{k=1}^{n_ij(y)} r_{ij}(y)_k+\alpha\right)}{\left(\prod_{k=1}^{n_{ij}(y)} r_{ij}(y)_k!\right)\Gamma(\alpha)}\beta^\alpha(n_{ij}(y)+\beta)^{-\left(\sum_{k=1}^{n_ij(y)} r_{ij}(y)_k+\alpha\right)} \tag{C.20}$$

Notice that the prefactor involving the gammas is essentially a multinomial. For concision let $\bar{r}_{ij}(y) = \sum_{k=1}^{n_{ij}(y)} r_{ij}(y)_k$ be the total runs batted by team i against team j in year y.

Then the negative log likelihood of α,β given data(Y) is (up to an additive constant):

$$\begin{aligned}-\log\left(\Pr\{\alpha,\beta|\text{data}(Y)\}\right) = &\sum_{y\in Y}\sum_{(ij)\in VE(y)}\log\left(\Gamma\left(\bar{r}_{ij}(y)+\alpha\right)\right)-\log\left(\Gamma(\alpha)\right)\\ &+\alpha\log(\beta)-\left(\bar{r}_{ij}(y)+\alpha\right)\log(n_{ij}(y)+\beta)+C\end{aligned} \tag{C.21}$$

where C is some additive constant that depends on the normalization.

Let $G(\alpha,\beta|\text{data}(Y)) = -\log\left(\Pr\{\alpha,\beta|\text{data}(Y)\}\right)$ without the added constant C. We can find the MLE for α,β by minimizing the $G(\alpha,\beta|\text{data}(Y))$. In practice this will be

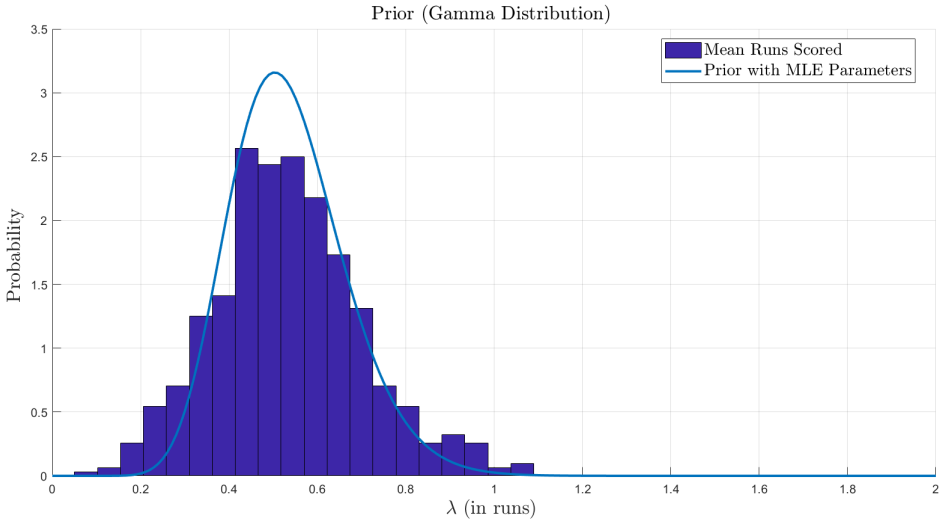

Figure C.2: Gamma prior distribution for baseball scoring rates in 2019 using MLE parameter estimates for α and β. The horizontal axis represents the expected number of runs per inning. The vertical axis represents the probability density.

done numerically however we can take some steps towards minimizing the negative log likelihood analytically. An example prior with MLE parameters is shown in Figure C.2.

First, the gradient of the negative log likelihood is:

$$\nabla G(\alpha, \beta | \text{data}(Y)) = -\sum_{y \in Y, (ij) \in VE(Y)} \begin{bmatrix} \psi(\bar{r}_{ij}(y) + \alpha) - \psi(\alpha) + \log(\beta) - \log(n_{ij}(y) + \beta) \\ (\alpha/\beta) - (\bar{r}_{ij}(y) + \alpha)/(n_{ij}(y) + \beta) \end{bmatrix} \tag{C.22}$$

where $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$ is the digamma function.

Settings the bottom partial to zero requires:

$$\frac{\alpha}{\beta} = \frac{1}{|\text{data}(Y)|} \sum_{y \in Y} \sum_{(ij) \in VE(y)} \frac{\bar{r}_{ij}(y) + \alpha}{n_{ij}(y) + \beta} \tag{C.23}$$

where $|\text{data}(Y)|$ is the total number of innings observed in the data. Notice that this requires that the ratio of total fictitious runs to fictitious innings equals the average over the data set

of the ratio of runs plus fictitious runs to innings plus fictitious innings. This shows that α/β must be consistent with the average runs per inning across the data set.

This is a linear equation in α_* so we can solve for $\alpha_*(\beta)$ such that at $\alpha_*(\beta), \beta$ the partial derivative with respect to β of G is zero. Rearranging:

$$\left(\frac{1}{\beta} - \frac{1}{|\text{data}(Y)|}\sum_{y\in Y}\sum_{(ij)\in VE(y)}\frac{1}{n_{ij}(y)+\beta}\right)\alpha = \frac{1}{|\text{data}(Y)|}\sum_{y\in Y}\sum_{(ij)\in VE(y)}\frac{\bar{r}_{ij}(y)}{n_{ij}(y)+\beta} \tag{C.24}$$

Therefore:

$$\alpha_*(\beta) = \frac{1}{\frac{|\text{data}(Y)|}{\beta} - \sum_{y\in Y}\sum_{(ij)\in VE(y)}\frac{1}{n_{ij}(y)+\beta}}\sum_{y\in Y}\sum_{(ij)\in VE(y)}\frac{\bar{r}_{ij}(y)}{n_{ij}(y)+\beta}. \tag{C.25}$$

This is the only solution where the partial with respect to β of the negative log likelihood is zero, so the minimum must lie on the curve $(\alpha_*(\beta), \beta)$. This turns the optimization problem for the prior parameters into a one-dimensional problem.

If the total number of runs observed is large and the fictitious number of runs observed is large we can approximate the digamma's in the partial with respect to α as $\log(\bar{r}_{ij}(y)+\alpha)$ and $\log(\alpha)$ (accurate to order $1/(\bar{r}_{ij}(y)+\alpha)$ and $1/\alpha$ respectively). Then, setting the top row to zero gives the approximate relation:

$$\frac{\alpha}{\beta} \simeq \left(\prod_{y\in Y}\prod_{(ij)\in VE(y)}\frac{\bar{r}_{ij}(y)+\alpha}{n_{ij}(y)+\beta}\right)^{1/|\text{data}(Y)|} \tag{C.26}$$

where this relation becomes true in the limit as α and $\bar{r}_{ij}(y)$ go to infinity. Therefore, when enough runs are observed the ratio of the MLE estimates of fictitious runs to fictitious innings must match both the arithmetic and geometric average of the ratio of (runs + fictitious runs) to (innings + fictitious innings).

We can also compute the Hessian of the negative log likelihood analytically. This is:

$$H(\alpha,\beta|\text{data}(Y)) = \sum_{y\in Y}\sum_{(ij)\in VE(Y)} \begin{bmatrix} \psi^{(1)}(\alpha) - \psi^{(1)}(\bar{r}_{ij}(y)+\alpha) & \frac{1}{n_{ij}(y)+\beta} - \frac{1}{\beta} \\ \frac{1}{n_{ij}(y)+\beta} - \frac{1}{\beta} & \frac{\alpha}{\beta^2} - \frac{\bar{r}_{ij}(y)+\alpha}{(n_{ij}(y)+\beta)^2} \end{bmatrix} \tag{C.27}$$

where $\psi^{(1)}(x)$ is the trigamma function.

Given the gradient and the Hessian analytically Newton's method can be used efficiently to solve for the minimizer of the negative log likelihood (since this is a two dimensional system we can even invert the Hessian analytically). Since $\alpha, \beta > 0$ the minimizer used is a constrained version of Newton's method. We can also easily compute the eigenvalue of the Hessian to check that it is positive definite (and hence that the negative log likelihood is convex).

The minimization is made easier by the fact that we have a good guess at the location of the minimizer. In general α/β should be close to the average runs scored per inning across the data set, so we pick initial estimates for the parameters so that their ratio matches the average runs scored per inning. In our experience setting β equal to the average number of innings played per pair, divided by two, gives a close initial estimate to the MLE parameters. This method has proved very fast and robust in all the examples tested. If desired we could even reduce the problem to a one dimensional problem by setting α equal to $\alpha_*(\beta)$ and performing the minimization in β alone.

C.4 Sampling Win Probabilities

We would like to be able to sample win probabilities from the posterior distribution of win probabilities given the data. This sampling can be done by first sampling prior parameters from the posterior distribution of prior parameters, then, given the prior parameters sam-

pled, sampling λ from the corresponding gamma distribution, and plugging the sampled λ's into the formula for computing win probabilities. Therefore the only new machinery we need in order to sample win probabilities from their posterior is a method for sampling the prior parameters.

This can be done efficiently using importance sampling. In importance sampling the samples are drawn from a proposal distribution that is designed to mimic the true distribution, but with a larger variance. Then, when averaging over the samples each sample is reweighted by its importance so that the weighted average mimics an average over the true distribution. If $p(x)$ is the true distribution, and $q(x)$ is the proposal distribution (which is supported everywhere $p(x)$ is supported) then using weights $w(x) = p(x)/q(x)$ and sampling $X \sim q$ gives:

$$\mathbb{E}_p[f(X)] \simeq \frac{\sum_j f(X_j)w(X_j)}{\sum_j w(X_j)}. \tag{C.28}$$

In our case this is an attractive sampling scheme since we do not need to know the normalizing constant of the true distribution (the weighted average is invariant under scaling by a constant), and we can use the Hessian to get a reasonable proposal distribution.

As a proposal distribution we use the Gaussian approximation to the posterior evaluated at the MLE parameters, with standard deviation scaled by a scale factor $s > 1$:

$$\mathcal{N}((\alpha_*, \beta_*), s^2 H(\alpha_*, \beta_*)^{-1}) \tag{C.29}$$

This ensures that the proposal distribution has the same shape as the posterior about the maximum likelihood parameters, but is spread more broadly, so samples rare events more efficiently. An example of the resulting samples and weights is show in Figure C.3

The Hessian typically has one large and one small singular value. The small singular value corresponds to a singular vector along the direction such that α/β is equal to the

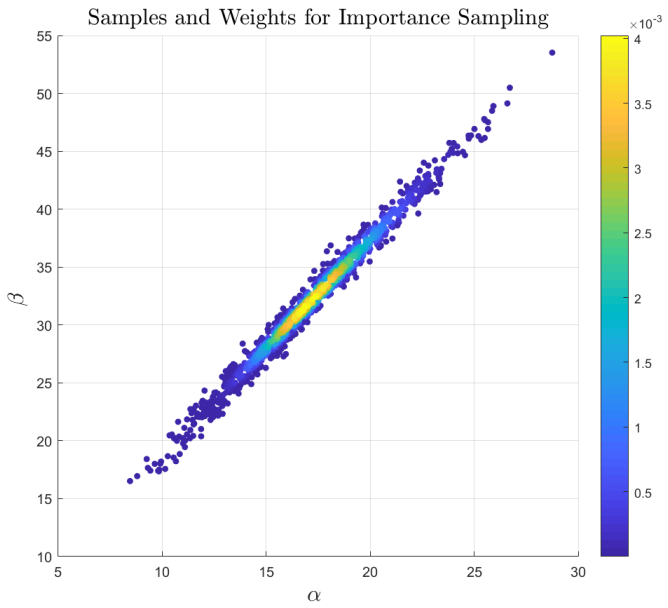

Figure C.3: Sampled prior parameters using importance sampling. Notice that the parameters are tightly distributed about the line such that α/β equals the average number of runs hit per inning in the season. The weights (importance) for each sample are color coded. Samples with blue weights are of low importance, while samples with yellow weights are of high importance.

average number of runs per inning. As a result the Gaussian approximation is tightly distributed perpendicular to this direction, but not in this direction. This anisotropy in the likelihood is thus captured by the proposal distribution.

In all of our trials we have found that setting $s = 2$ gives a proposal distribution such that the weights decay to zero in all directions away from the maximum likelihood parameter values.

Appendix D

Building an Optimal Spanning Tree

The steady state of a Markov process in all three strongly forced limits ($\beta \rightarrow \infty$) is described by the work evaluated over an optimal spanning tree. If β goes to infinity, and the conductances are held fixed, then the work is evaluated against the edge flow f, and for each node we seek the directed spanning tree oriented towards that node that maximizes the work exerted by the edge flow along the path from each leaf of the tree back to the node of interest. This is, in effect, the total energy exchanged with the reservoir along an ensemble of relaxation trajectories. The ratio of the steady state probabilities is given by evaluating the difference in these trees. In the strong rotational and near deterministic limits the work is evaluated against $\Delta f_{ij} = f_{ij} - \max_{k \in \mathcal{N}_i}\{f_{ik}\}$, and the optimal spanning tree is the collection of paths which minimize the work exerted by the process to move along activation trajectories away from a stable cycle or pair of nodes. In each case the value of the work evaluated over these trees defines a quasipotential like object which converges to the effective potential (log of the steady state) in the limit.

To compute either of these network quasipotentials one must first construct an optimal

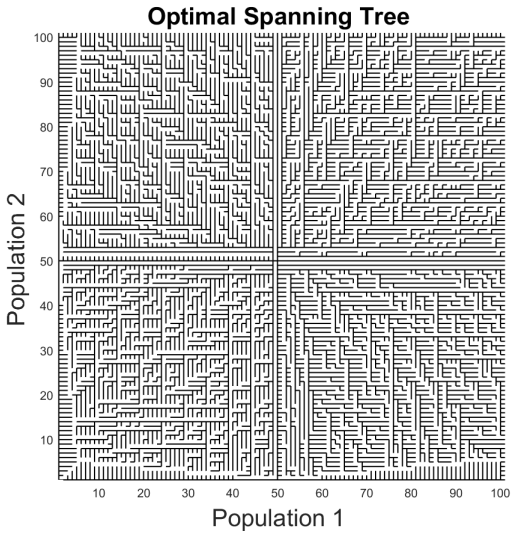

Figure D.1: An example optimal spanning tree for a two dimensional lattice with $x_0 = [50, 50]$ and with edge rates set to approximate an OU process with constant noise.

spanning tree. A simple algorithm for constructing optimal spanning trees is presented in this section. An example minimizing the work against f on a lattice with transition rates chosen to approximate an OU process is shown in Figure D.1.

A network quasipotential is a function on the nodes that is defined by a functional S. The functional acts on the space of paths and returns a real number. The functional can be thought of as the action or work associated with specific paths.

In order to ensure that the optimization problem is tractable we require that the functional satisfies certain requirements. The first requirement ensures that we can compute the value of the functional one step of the path at a time. The second ensures that lengthening a path always increases or decreases the value of the functional:

1. Let y be a path from a to b. Let c be a neighbor of b. Let y' be the path from a to c that follows y to b then the edge from b to c. We require $S(y') = H[S(y), b, c]$ where $H[s, b, c]$ is an update function that depends exclusively on the value s and the edge b, c.

2. Monotonicity: either $H[S(y), b, c] \leq S_1(y)$ for all $S(y), b, c$ or $H[S(y), b, c] \geq S(y)$ for all $S(y), b, c$. Moreover, if $s > t$ then $H[s, b, c] > H[t, b, c]$ for all s, t, b, c.

These two conditions are satisfied by a variety of natural functionals. Path integrals over f always satisfy the first condition, and path integrals over $\min\{f, 0\}$ or $\max\{f, 0\}$ always satisfy the second. Path integrals against Δf satisfy both. We have seen these path integrals arise repeatedly when considering the work over paths.

Alternatively, given a skeleton trajectory X consisting of n nodes on a path from a to b, the probability of walking the trajectory is:

$$S(X) = \prod_{j=1} \frac{l_{x_{j+1},x_j}}{|l_{x_j,x_j}|} \tag{D.1}$$

Where x_j is the j^{th} node in the trajectory.

This satisfies the first condition since:

$$S([x_1, x_2, ...x_n]) = \frac{l_{x_n,x_{n-1}}}{|l_{x_{n-1},x_{n-1}}|} S([x_1, x_2, ...x_{n-1}]) = H\left[S([x_1, x_2, ...x_{n-1}]), x_{n-1}, x_n\right]. \tag{D.2}$$

It is also monotonic, since $l_{x_{j+1},x_j} \leq |l_{x_j,x_j}|$. The probability of walking a specific trajectory never increases when we add an edge to the trajectory. In a strong forcing limit the log of this probability converges to the work evaluated against Δf, the difference between f_{ij} and the largest flow leaving node i.

Next we define the optimal trajectory (or set of optimal trajectories) from a to b given S to be the trajectory (or set of trajectories) that minimizes S if S is monotonically increasing, or maximizes S if S is monotonically decreasing. The monotonicity condition is essential since it ensures that optimal trajectories are never infinitely long , never contain loops, and never contain any edges in both the forward and backward direction.

Fix an initial node x_0. An optimal spanning tree, denoted $\mathcal{T}[S_1, x_0]$, is a spanning tree of the network $\mathcal{G}$ such that each path in the spanning tree is an optimal path with respect to S.

It is obvious that any set of optimal paths from x_0 to every other node spans the network. It is not obvious to see that these paths always form a tree. In order to form a tree they need to be self consistent. That is, the optimal path $X = [x_0, x_1, ...x_n]$ is equivalent when truncated to $x_j, j < n$, to the optimal path from x_0 to x_j. It is sufficient to show this equivalence for $j = n - 1$, since equivalence for all j follows by induction.

Consider a node $x_n \neq x_0$. There always exists at least one optimal path X from x_0 to x_n. Let x_{n-1} be the node preceding x_n in the optimal trajectory X, and $X' = [x_0, x_1, ...x_{n-1}]$. Let Y' denote an optimal trajectory from x_0 to x_{n-1}. If Y' can always be chosen so that it is equivalent to X' then it is always possible to pick a set of optimal paths that form an spanning tree. To show that X' is always equivalent to some optimal path Y' we proceed by contradiction.

By the first requirement on S:

$$S[X] = H[S[X'], x_{n-1}, x_n].$$

Let Y be the path from x_0 to x_n that follows Y' to x_{n-1} then moves directly from x_{n-1} to x_n. Then, by definition:

$$S[Y] = H[S[Y'], x_{n-1}, x_n].$$

Suppose X' is not an optimal path from x_0 to x_{n-1}. Then $S[X'] > S[Y']$ for any

optimal Y. Then by the monotonicity requirement:

$$S[X] = H[S[X'], x_{n-1}, x_n] > H[S[Y'], x_{n-1}, x_n] = S[Y].$$

But Y is a path from x_0 to x_n so if $S[Y] < S[X]$ then X is not an optimal trajectory. Therefore, if X is an optimal trajectory from x_0 to x_n then X' must be an optimal trajectory from x_0 to x_{n-1}. It follows by induction that any truncation of the an optimal trajectory X is itself an optimal trajectory.

So, given an initial node x_0 and a functional S that satisfies the first and second requirement it is always possible to construct a spanning tree $\mathcal{T}[S, x_0]$ such that every path in the tree from x_0 to x is an optimal path with respect to S.

Next we need an algorithm for constructing an optimal spanning tree given an action functional S and an initial node x_0. Thankfully, since S is monotonic, and can be computed one edge at a time, this algorithm is both easy to construct and surprisingly efficient. The key idea when constructing an optimal spanning tree is that a truncation of an optimal path is itself an optimal path. It follows that we can build the tree outwards by considering all optimal paths of a set length, starting with length one, then length two, and so on.

The process is initialized by indexing all the nodes and picking x_0. Without loss of generality assume that S is monotonically increasing as paths grow longer. Define a vector u with as many entries as there are nodes. Set $u(x_0) = 0$. When considering all paths length k or less the i^{th} entry of p will correspond to the optimal value of S over all paths from x_0 to the i^{th} node with k or fewer edges.

Any spanning tree is completely specified by listing all of the edges of the spanning tree. Since there are $V - 1$ edges we list one edge for every node that is not the root x_0. A convenient way to list the edges is to orient each edge so that it points out from x_0, that is

from the node closer to x_0 in the tree to the node farther from x_0 in the tree. Then, every node except x_0 has exactly one “parent" node, corresponding to the node one step closer to x_0. Then the entire tree can be stored by a vector with V entries, whose i^{th} entry is an index corresponding to the parent of i. In this context the node i is the “child" of the node j. Keeping with the terminology the vector that lists all the parents of each child is the genealogy vector.

We build the tree by starting with all paths length less than or equal to one, then less than or equal to two, and so on. Let k represent the length of the largest possible path considered on the k^{th} iteration. At the end of the k^{th} step the set of all the nodes who are a distance k from x_0 in the current tree are the k^{th} generation.

Suppose that, at the end of step k we have a optimal spanning tree of all the nodes within k steps of x_0 using paths of length k or less. This is stored as the k^{th} iteration of the genealogy vector. To find the optimal tree of all paths from x_0 with length $k+1$ or less we need only consider paths that extend paths length k. That is, the possible children in the $k+1^{st}$ step must all be neighbors of the nodes in the k^{th} generation. So the list of possible parents of the $k+1^{st}$ step is the list of children of the k^{th} step.

Loop over the nodes in the k^{th} generation. Each node is possibly a new parent. Let the index j refer to the parent. Then loop over all the possible children of each parent. Note that this includes all neighbors of j except for j's parent. The index i will refer to a specific possible child. Since each parent j is included in the spanning tree of paths less than length $k+1$ we know the optimal value of S over all paths length less than $k+1$ from x_0 to j. This was stored as u_j. Now, given a specific parent j, and given a specific child i, the optimal spanning tree with paths of length less than or equal to k includes a specific path with length $k+1$ from the node x_0 to the node j, then to the node i. Denote this path X. To find the value of S along this path we use the edge based update rule $S[X] = H[u_j, j, i]$.

If the child i has not yet been considered then we automatically add it to the list of new children, with parent j. If the child has already been considered we compare $S[X]$ to u_i. If $S[X] < u_i$ then the path X is preferable to the path we had previously considered from x_0 to i. In that case, switch the parent of node i to node j, and let $u_i = H[u_j, j, i]$. Repeating this process for every possible child of every possible parent produces a new genealogy corresponding to the tree of optimal paths length $k + 1$ or less. Every node whose parentage changed in the $k + 1^{st}$ step is stored as a possible parent for the next step.

This process is repeated until no new nodes are added. The process necessarily stops in finite time since the monotonicity of S ensures the optimal paths never form loops, and the longest non-looping path through a network with V nodes has $V - 1$ edges. Therefore, at most, the process runs for $V - 1$ stages.

It is worth noting that this algorithm is entirely general, so can compute optimal spanning trees for any appropriate choice of S that is monotonic and can be updated one edge at a time. It is also worth noting that if there are multiple optimal spanning trees for a given S, x_0 then this algorithm only returns one of the multiple possible trees. To return all of the trees replace each entry of the genealogy vector with a list of all the possible parents of the given node corresponding to optimal spanning trees. That is, in the special case that we find $S_1[X] = u_i$ simply add the node j to the list of possible parents of i, instead of clearing the list and using j alone. Since the quasipotential is the value of S over the optimal trajectories, if there are multiple optimal spanning trees all must have the same quasipotential, so it suffices to find one optimal spanning tree.

An example optimal spanning tree and quasipotential with S set to the likelihood of skeleton trajectories is shown in Figure 7.11 along with the associated network quasipotential. The transition rates are chosen to approximate an OU process, which has linear drift, so the corresponding potential is approximately quadratic.

Part VII

Bibliography

Bibliography

[1] James Clerk Maxwell. On Faraday's lines of force. Printed at the University Press by CJ Clay. printer to the University, 1856.

[2] Hermann von Helmholtz. On integrals of the hydrodynamic equations that correspond to vortex motions. International Journal of Fusion Energy, 1(3-4):41–68, 1978.

[3] Sophia L Kalpazidou. Cycle representations of Markov processes, volume 28. Springer Science & Business Media, 2007.

[4] Hong Qian. Nonequilibrium steady-state circulation and the heat dissipation functional. Physical Review E, 64(2):022101, 2001.

[5] Jürgen Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. Reviews of Modern physics, 48(4):571, 1976.

[6] Terrell L Hill and Yi-Der Chen. Stochastics of cycle completions (fluxes) in biochemical kinetic diagrams. Proceedings of the National Academy of Sciences, 72(4):1291–1295, 1975.

[7] Terrell Hill. Free energy transduction in biology: the steady-state kinetic and thermodynamic formalism. Elsevier, 2012.

[8] Harsh Bhatia, Gregory Norgard, Valerio Pascucci, and Peer-Timo Bremer. The Helmholtz-Hodge decomposition—a survey. IEEE Transactions on visualization and computer graphics, 19(8):1386–1404, 2012.

[9] M Akram and V Michel. Regularisation of the Helmholtz decomposition and its application to geomagnetic field modelling. GEM-International Journal on Geomathematics, 1(1):101–120, 2010.

[10] Hengzhen Gao, Mrinal K Mandal, Gencheng Guo, and Jianwei Wan. Singular point detection using the discrete Hodge Helmholtz decomposition in fingerprint images. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, pages 1094–1097. IEEE, 2010.

[11] Qinghong Guo, Mrinal K Mandal, Gang Liu, and Katherine M Kavanagh. Cardiac video analysis using Hodge–Helmholtz field decomposition. Computers in Biology and Medicine, 36(1):1–20, 2006.

[12] Nagi N Mansour, Dali Georgobiani, Alexander G Kosovichev, Robert F Stein, and Åke Nordlund. Decomposition of turbulent velocity fields in numerical simulations of solar convection.

[13] Nagi N Mansour, AG Kosovichev, Dali Georgobiani, Alan Wray, and Mark Miesch. Turbulence convection and oscillations in the sun. In SOHO 14 Helio-and Asteroseismology: Towards a Golden Future, volume 559, page 164, 2004.

[14] Yoshihiko Mochizuki and Atsushi Imiya. Spatial reasoning for robot navigation using the Helmholtz-Hodge decomposition of omnidirectional optical flow. In 2009 24th International Conference Image and Vision Computing New Zealand, pages 1–6. IEEE, 2009.

[15] Lek-Heng Lim. Hodge Laplacians on graphs. arXiv preprint arXiv:1507.05379, 2015.

[16] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. Mathematical Programming, 127(1):203–244, 2011.

[17] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Learning to rank with combinatorial hodge theory. stat, 1050:7, 2008.

[18] Ozan Candogan, Ishai Menache, Asuman Ozdaglar, and Pablo A Parrilo. Flows and decompositions of games: Harmonic and potential games. Mathematics of Operations Research, 36(3):474–503, 2011.

[19] Jürgen Schnakenberg. Thermodynamic network analysis of biological systems. Springer Science & Business Media, 2012.

[20] Hong Qian, Min Qian, and Xiang Tang. Thermodynamics of the general diffusion process: time-reversibility and entropy production. Journal of statistical physics, 107(5-6):1129–1141, 2002.

[21] Hong Qian. Thermodynamics of the general diffusion process: Equilibrium supercurrent and nonequilibrium driven circulation with dissipation. The European Physical Journal Special Topics, 224(5):781–799, 2015.

[22] Hong Qian. The mathematical theory of molecular motor movement and chemomechanical energy transduction. Journal of Mathematical Chemistry, 27(3):219–234, 2000.

[23] Ben C Nolting and Karen C Abbott. Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. Ecology, 2016.

[24] Paul C Bressloff. Stochastic processes in cell biology, volume 41. Springer, 2014.

[25] AD Wentzell and MI Freidlin. Small random perturbations of dynamical systems (russian), uspehi mat. Nauk, 25:3–55, 1970.

[26] Gilbert Strang. The fundamental theorem of linear algebra. The American Mathematical Monthly, 100(9):848–855, 1993.

[27] Donald H Kobe. Helmholtz's theorem revisited. American Journal of Physics, 54(6):552–554, 1986.

[28] Peter G Doyle and J Laurie Snell. Random walks and electric networks, volume 22. American Mathematical Soc., 1984.

[29] Otto Blumenthal. About the decomposition of infinite vector fields. Mathematical Annals, 61(2):235–250, 1905.

[30] Harsh Bhatia, Valerio Pascucci, and Peer-Timo Bremer. The natural Helmholtz-Hodge decomposition for open-boundary flow analysis. IEEE transactions on visualization and computer graphics, 20(11):1566–1578, 2014.

[31] Filippo Maria Denaro. On the application of the Helmholtz–Hodge decomposition in projection methods for incompressible flows with general boundary conditions. International Journal for Numerical Methods in Fluids, 43(1):43–69, 2003.

[32] John Hopcroft and Robert Tarjan. Efficient planarity testing. Journal of the ACM (JACM), 21(4):549–568, 1974.

[33] Telikepalli Kavitha, Christian Liebchen, Kurt Mehlhorn, Dimitrios Michail, Romeo Rizzi, Torsten Ueckerdt, and Katharina A Zweig. Cycle bases in graphs characterization, algorithms, complexity, and applications. Computer Science Review, 3(4):199–243, 2009.

[34] Béla Bollobás. Modern graph theory, volume 184. Springer Science & Business Media, 2013.

[35] Christian Liebchen and Romeo Rizzi. Classes of cycle bases. Discrete Applied Mathematics, 155(3):337–355, 2007.

[36] David Hartvigsen and Russell Mardon. The all-pairs min cut problem and the minimum cycle basis problem on planar graphs. SIAM Journal on Discrete Mathematics, 7(3):403–418, 1994.

[37] Ugur Dog¯ Rusöz and MS Krishnamoorthy. Enumerating all cycles of a planar graph. International Journal of Parallel, Emergent and Distributed Systems, 10(1-2):21–36, 1996.

[38] G Kirchhoff. About the resolution of the equations to which one is led in the investigation of the linear distribution of galvanic currents. English transl., Trans. IRE CT-5, pages 4–7, 1958.

[39] Alexander Golynski and Joseph D Horton. A polynomial time algorithm to find the minimum cycle basis of a regular matroid. In Scandinavian Workshop on Algorithm Theory, pages 200–209. Springer, 2002.

[40] Richard Hammack. Minimum cycle bases of direct products of complete graphs. Information Processing Letters, 102(5):214–218, 2007.

[41] Marc Hellmuth, Josef Leydold, and Peter F Stadler. Convex cycle bases. Ars Mathematica Contemporanea, 7(1):123–140, 2014.

[42] Joseph Douglas Horton. A polynomial-time algorithm to find the shortest cycle basis of a graph. SIAM Journal on Computing, 16(2):358–366, 1987.

[43] Telikepalli Kavitha and Kurt Mehlhorn. A polynomial time algorithm for minimum cycle basis in directed graphs. In Annual Symposium on Theoretical Aspects of Computer Science, pages 654–665. Springer, 2005.

[44] Wilfried Imrich and Peter F Stadler. Minimum cycle bases of product graphs. Australasian Journal of Combinatorics, 26:233–244, 2002.

[45] Saunders Mac Lane. A combinatorial condition for planar graphs. Seminarium Matemat., 1936.

[46] Bernhard Altaner, Stefan Grosskinsky, Stephan Herminghaus, Lukas Katthän, Marc Timme, and Jürgen Vollmer. Network representations of nonequilibrium steady states: Cycle decompositions, symmetries, and dominant paths. Physical Review E, 85(4):041133, 2012.

[47] Edward C Kirby, Roger B Mallion, Paul Polla, and Pawel J Skrzynski. What Kirchhoff actually did concerning spanning trees in electrical networks and its relationship to modern graph-theoretical work. Croatica Chemica Acta, 89(4):1–16, 2016.

[48] Yehuda Koren. Drawing graphs by eigenvectors: theory and practice. Computers & Mathematics with Applications, 49(11-12):1867–1888, 2005.

[49] Daniel A Spielman. Spectral graph theory and its applications. In 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), pages 29–38. IEEE, 2007.

[50] Daniel Spielman. Spectral graph theory. In Combinatorial scientific computing, number 18. Citeseer, 2012.

[51] Lloyd N Trefethen and David Bau III. Numerical linear algebra, volume 50. Siam, 1997.

[52] Emmanuel J Candès, Yaniv Plan, et al. Near-ideal model selection by l_1 minimization. The Annals of Statistics, 37(5A):2145–2177, 2009.

[53] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. Proceedings of the National Academy of Sciences, 100(5):2197–2202, 2003.

[54] Daniela Calvetti, Erkki Somersalo, and A Strang. Hierachical Bayesian models and sparsity: l_2-magic. Inverse Problems, 35(3):035003, 2019.

[55] Daniela Calvetti, Monica Pragliola, Erkki Somersalo, and Alexander Strang. Sparse reconstructions from few noisy data: analysis of hierarchical bayesian models with generalized gamma hyperpriors. Inverse Problems, 36(2):025010, 2020.

[56] Geoffrey Grimmett, Geoffrey R Grimmett, David Stirzaker, et al. Probability and random processes. Oxford university press, 2001.

[57] Alexandre Joel Chorin, Jerrold E Marsden, and Jerrold E Marsden. A mathematical introduction to fluid mechanics, volume 3. Springer, 1990.

[58] John E. Hopcroft and Robert Endre Tarjan. Efficient algorithms for graph manipulation [H] (Algorithm 447). Commun. ACM, 16(6):372–378, 1973.

[59] Robert Tarjan. Depth-first search and linear graph algorithms. SIAM journal on computing, 1(2):146–160, 1972.

[60] Sheshayya A Choudum and N Priya. Tenacity of complete graph products and grids. Networks: An International Journal, 34(3):192–196, 1999.

[61] A Kaveh and H Rahami. An efficient method for decomposition of regular structures using graph products. International Journal for Numerical Methods in Engineering, 61(11):1797–1808, 2004.

[62] A Kaveh, M Nikbakht, and H Rahami. Improved group theoretic method using graph products for the analysis of symmetric-regular structures. Acta mechanica, 210(3-4):265–289, 2010.

[63] Stefan Janicke, Christian Heine, Marc Hellmuth, Peter F Stadler, and Gerik Scheuermann. Visualization of graph products. IEEE Transactions on Visualization and Computer Graphics, 16(6):1082–1089, 2010.

[64] Richard Hammack, Wilfried Imrich, and Sandi Klavžar. Handbook of product graphs. CRC press, 2011.

[65] A Kaveh and R Mirzaie. Minimal cycle basis of graph products for the force method of frame analysis. Communications in numerical methods in engineering, 24(8):653–669, 2008.

[66] Gert Sabidussi. Graph multiplication. mathematical journal, 72(1):446–457, 1959.

[67] Wilfried Imrich and Janez Žerovnik. Factoring Cartesian-product graphs. Journal of Graph Theory, 18(6):557–567, 1994.

[68] Sandi Klavžar and Iztok Peterin. Characterizing subgraphs of Hamming graphs. Journal of graph theory, 49(4):302–312, 2005.

[69] Charles F Van Loan. The ubiquitous Kronecker product. Journal of computational and applied mathematics, 123(1-2):85–100, 2000.

[70] Henri J Nussbaumer. The Fast Fourier Transform. In Fast Fourier Transform and Convolution Algorithms, pages 80–111. Springer, 1981.

[71] Deena R Schmidt, Roberto F Galán, and Peter J Thomas. Stochastic shielding and edge importance for Markov chains with timescale separation. PLoS computational biology, 14(6):e1006206, 2018.

[72] Peter Kirrinnis. Fast algorithms for the Sylvester equation $AX - XB^{\intercal} = C$. Theoretical Computer Science, 259(1-2):623–638, 2001.

[73] Fred W Dorr. The direct solution of the discrete Poisson equation on a rectangle. SIAM review, 12(2):248–263, 1970.

[74] Gene Golub and Charles Van Loan. Matrix computations. Matrix, 1000(13):09, 1996.

[75] Crispin W Gardiner et al. Handbook of stochastic methods, volume 3. Springer Berlin, 1985.

[76] Randall J LeVeque. Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems, volume 98. Siam, 2007.

[77] William K Pratt, Julius Kane, and Harry C Andrews. Hadamard transform image coding. Proceedings of the IEEE, 57(1):58–68, 1969.

[78] M Lee and Mostafa Kaveh. Fast Hadamard transform based on a simple matrix factorization. IEEE transactions on acoustics, speech, and signal processing, 34(6):1666–1667, 1986.

[79] Henry O. Kunz. On the equivalence between one-dimensional discrete Walsh-Hadamard and multidimensional discrete Fourier transforms. IEEE Transactions on Computers, (3):267–268, 1979.

[80] Tom Beer. Walsh transforms. American Journal of Physics, 49(5):466–472, 1981.

[81] A Kaveh and GR Roosta. Revised greedy algorithm for formation of a minimal cycle basis of a graph. Communications in numerical methods in engineering, 10(7):523–530, 1994.

[82] M Jaradat. Minimal cycle bases of the lexicographic product of graphs. Discussiones Mathematicae Graph Theory, 28(2):229–247, 2008.

[83] Mark EJ Newman et al. Random graphs as models of networks. Handbook of graphs and networks, 1:35–68, 2003.

[84] Fan Chung and Linyuan Lu. The diameter of sparse random graphs. Advances in Applied Mathematics, 26(4):257–279, 2001.

[85] Fan Chung and Linyuan Lu. The average distance in a random graph with given expected degrees. Internet Mathematics, 1(1):91–113, 2004.

[86] Béla Bollobás. The diameter of random graphs. Transactions of the American Mathematical Society, 267(1):41–52, 1981.

[87] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. Combinatorica, 24(1):5–34, 2004.

[88] Béla Bollobás and W Fernandez De La Vega. The diameter of random regular graphs. Combinatorica, 2(2):125–134, 1982.

[89] B Bollobás. Random graphs Academic Press. New York, 1985.

[90] Edoardo Amaldi, Claudio Iuliano, Tomasz Jurkiewicz, Kurt Mehlhorn, and Romeo Rizzi. Breaking the $\mathcal{O}(m^2n)$ barrier for minimum cycle bases. In European Symposium on Algorithms, pages 301–312. Springer, 2009.

[91] A Lempel, S Even, and I Cederbaum. An algorithm for planarity testing of graphs. theory of graphs: International symposium: Rome, July, 1966, P. Rosenstiehl, Ed, 1967.

[92] Lee F Mondshein. Combinatorial ordering and the geometric embedding of graphs. Technical report, Massachusetts Institute of Technology Lexington Lincoln Lab, 1971.

[93] Prabhaker Mateti and Narsingh Deo. On algorithms for enumerating all circuits of a graph. SIAM Journal on Computing, 5(1):90–99, 1976.

[94] Robert A Laird and Brandon S Schamp. Competitive intransitivity promotes species coexistence. The American Naturalist, 168(2):182–193, 2006.

[95] Daizaburo Shizuka and David B McDonald. A social network perspective on measurements of dominance hierarchies. Animal Behaviour, 83(4):925–934, 2012.

[96] Sándor Bozóki, László Csató, and József Temesi. An application of incomplete pairwise comparison matrices for ranking top tennis players. European Journal of Operational Research, 248(1):211–218, 2016.

[97] Maurice G Kendall and B Babington Smith. On the method of paired comparisons. Biometrika, 31(3/4):324–345, 1940.

[98] William V Gehrlein. Condorcet's paradox. Springer, 2006.

[99] Robert M May and Warren J Leonard. Nonlinear aspects of competition between three species. SIAM journal on applied mathematics, 29(2):243–253, 1975.

[100] Peter S Petraitis. Competitive networks and measures of intransitivity. The American Naturalist, 114(6):921–925, 1979.

[101] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model. Physical Review E, 74(5):051907, 2006.

[102] Benjamin Kerr, Margaret A Riley, Marcus W Feldman, and Brendan JM Bohannan. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. Nature, 418(6894):171–174, 2002.

[103] Barry Sinervo and Curt M Lively. The rock–paper–scissors game and the evolution of alternative male strategies. Nature, 380(6571):240–243, 1996.

[104] Patrick Slater. Inconsistencies in a schedule of paired comparisons. Biometrika, 48(3/4):303–312, 1961.

[105] Tobias Reichenbach and Erwin Frey. Instability of spatial patterns and its ambiguous impact on species diversity. Physical review letters, 101(5):058102, 2008.

[106] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. Nature, 448(7157):1046–1049, 2007.

[107] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. Noise and correlations in a spatial population model with cyclic competition. Physical review letters, 99(23):238105, 2007.

[108] Chi Xue and Nigel Goldenfeld. Coevolution maintains diversity in the stochastic "Kill the Winner" model. Physical review letters, 119(26):268101, 2017.

[109] JBC Jackson and LEO Buss. Alleopathy and spatial competition among coral reef invertebrates. Proceedings of the National Academy of Sciences, 72(12):5160–5163, 1975.

[110] Richard A Lankau and Sharon Y Strauss. Mutual feedbacks maintain both genetic and species diversity in a plant community. science, 317(5844):1561–1563, 2007.

[111] Richard A Lankau, Emily Wheeler, Alison E Bennett, and Sharon Y Strauss. Plant–soil feedbacks contribute to an intransitive competitive network that promotes both genetic and species diversity. Journal of Ecology, 99(1):176–185, 2011.

[112] Oscar Godoy, Daniel B Stouffer, Nathan JB Kraft, and Jonathan M Levine. Intransitivity is infrequent and fails to promote annual plant coexistence without pairwise niche differences. Ecology, 98(5):1193–1200, 2017.

[113] Santiago Soliveres, Fernando T Maestre, Werner Ulrich, Peter Manning, Steffen Boch, Matthew A Bowker, Daniel Prati, Manuel Delgado-Baquerizo, José L Quero, Ingo Schöning, et al. Intransitive competition is widespread in plant communities and maintains their species richness. Ecology letters, 18(8):790–798, 2015.

[114] Werner Ulrich, Santiago Soliveres, Wojciech Kryszewski, Fernando T Maestre, and Nicholas J Gotelli. Matrix models for quantifying competitive intransitivity from species abundance data. Oikos, 123(9):1057–1070, 2014.

[115] Pierre Charbit, Stéphan Thomassé, and Anders Yeo. The minimum feedback arc set problem is NP-hard for tournaments. Combinatorics, Probability and Computing, 16(1):1–4, 2007.

[116] Peter Eades, Xuemin Lin, and William F Smyth. A fast and effective heuristic for the feedback arc set problem. Information Processing Letters, 47(6):319–323, 1993.

[117] Hyman G Landau. On dominance relations and the structure of animal societies: I. Effect of inherent characteristics. The bulletin of mathematical biophysics, 13(1):1–19, 1951.

[118] Devi M Stuart-Fox, David Firth, Adnan Moussalli, and Martin J Whiting. Multiple signals in chameleon contests: designing and analysing animal contests as a tournament. Animal Behaviour, 71(6):1263–1271, 2006.

[119] Michael P Haley, Charles J Deutsch, and Burney J Le Boeuf. Size, dominance and copulatory success in male northern elephant seals, mirounga angustirostris. Animal Behaviour, 48(6):1249–1260, 1994.

[120] Michael Lewis. Moneyball: The art of winning an unfair game. WW Norton & Company, 2004.

[121] C Soto Valero. Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods. International Journal of Computer Science in Sport, 15(2):91–112, 2016.

[122] Mark E Glickman. Parameter estimation in large dynamic paired comparison experiments. Journal of the Royal Statistical Society: Series C (Applied Statistics), 48(3):377–394, 1999.

[123] Amy N Langville and Carl D Meyer. Who's# 1?: the science of rating and ranking. Princeton University Press, 2012.

[124] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. Computer networks, 56(18):3825–3833, 2012.

[125] Ray Stefani. The methodology of officially recognized international sports rating systems. Journal of Quantitative Analysis in Sports, 7(4), 2011.

[126] M Kwiesielewicz. The logarithmic least squares and the generalized pseudoinverse in estimating ratios. European Journal of Operational Research, 93(3):611–619, 1996.

[127] Miroslaw Kwiesielewicz and Ewa Van Uden. Ranking decision variants by subjective paired comparisons in cases with incomplete data. In International Conference on Computational Science and Its Applications, pages 208–215. Springer, 2003.

[128] Wesley Colley. Colley's bias free college football ranking method, 2002.

[129] James P Keener. The Perron–Frobenius theorem and the ranking of football teams. SIAM review, 35(1):80–93, 1993.

[130] Kenneth Massey. Statistical models applied to the rating of sports teams. Bluefield College, 1997.

[131] Raymond T Stefani. Football and basketball predictions using least squares. IEEE Transactions on systems, man, and cybernetics, 7(2):117–21, 1977.

[132] Raymond T Stefani. Improved least squares football, basketball, and soccer predictions. IEEE transactions on systems, man, and cybernetics, 10(2):116–123, 1980.

[133] Michael C Appleby. The probability of linearity in hierarchies. Animal Behaviour, 31(2):600–608, 1983.

[134] John Bartholdi, Craig A Tovey, and Michael A Trick. Voting schemes for which it can be difficult to tell who won the election. Social Choice and welfare, 6(2):157–165, 1989.

[135] Ulle Endriss and Ronald de Haan. Complexity of the winner determination problem in judgment aggregation: Kemeny, Slater, Tideman, Young. (downloaded from https://eprints.illc.uva.nl/534/), 2016.

[136] Edith Hemaspaandra, Holger Spakowski, and Jörg Vogel. The complexity of Kemeny elections. Theoretical Computer Science, 349(3):382–391, 2005.

[137] David Aldous et al. Elo ratings and the sports model: A neglected topic in applied probability? Statistical Science, 32(4):616–629, 2017.

[138] Lars Magnus Hvattum and Halvard Arntzen. Using Elo ratings for match result prediction in association football. International Journal of forecasting, 26(3):460–470, 2010.

[139] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika, 39(3/4):324–345, 1952.

[140] Ralph Allan Bradley. Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. Biometrics, 10(3):375–390, 1954.

[141] George Rabinowitz. Some comments on measuring world influence. Journal of Peace Science, 2(1):49–55, 1976.

[142] Yannis Sismanis. How I won the “Chess Ratings – Elo vs the Rest of the World" competition. arXiv preprint arXiv:1012.4571, 2010.

[143] Han De Vries. An improved test of linearity in dominance hierarchies containing unknown or tied relationships. Animal Behaviour, 50(5):1375–1389, 1995.

[144] Paul W Holland and Samuel Leinhardt. Local structure in social networks. Sociological methodology, 7:1–45, 1976.

[145] Peyton Young. Optimal voting rules. Journal of Economic Perspectives, 9(1):51–64, 1995.

[146] Vincent Conitzer, Andrew Davenport, and Jayant Kalagnanam. Improved bounds for computing Kemeny rankings. In AAAI, volume 6, pages 620–626, 2006.

[147] Jan Lasek, Zoltán Szlávik, and Sandjai Bhulai. The predictive power of ranking systems in association football. International Journal of Applied Pattern Recognition, 1(1):27–46, 2013.

[148] Ian McHale and Alex Morton. A Bradley-Terry type model for forecasting tennis match results. International Journal of Forecasting, 27(2):619–630, 2011.

[149] Stephanie Ann Kovalchik. Searching for the GOAT of tennis win prediction. Journal of Quantitative Analysis in Sports, 12(3):127–138, 2016.

[150] Giulia Galbiati. On optimum cycle bases. Electronic Notes in Discrete Mathematics, 10:113–116, 2001.

[151] Marcus Frean and Edward R Abraham. Rock–scissors–paper and the survival of the weakest. Proceedings of the Royal Society of London. Series B: Biological Sciences, 268(1474):1323–1327, 2001.

[152] Craig R Johnson and Ingrid Seinen. Selection for restraint in competitive ability in spatial competition systems. Proceedings of the Royal Society of London. Series B: Biological Sciences, 269(1492):655–663, 2002.

[153] Mike Ozanian and Kurt Badenhausen. Baseball team values 2019. Forbes, Apr 2019.

[154] Michael Luca and Jonathan Smith. Salience in quality disclosure: Evidence from the US News college rankings. Journal of Economics & Management Strategy, 22(1):58–77, 2013.

[155] Patricia M McDonough, Anthony Lising, Antonio Marybeth Walpole, and Leonor Xochitl Perez. College rankings: Democratized college knowledge for whom? Research in higher education, 39(5):513–537, 1998.

[156] James Monks and Ronald G Ehrenberg. US News & World Report's college rankings: Why they do matter. Change: The Magazine of Higher Learning, 31(6):42–51, 1999.

[157] Kurt Bryan and Tanya Leise. The $ 25,000,000,000 eigenvector: The linear algebra behind Google. SIAM review, 48(3):569–581, 2006.

[158] Robert M Bell, Yehuda Koren, and Chris Volinsky. The Bellkor solution to the Netflix prize. KorBell Team's Report to Netflix, 2007.

[159] Robert M Bell and Yehuda Koren. Lessons from the Netflix prize challenge. Acm Sigkdd Explorations Newsletter, 9(2):75–79, 2007.

[160] Robert M Bell, Yehuda Koren, and Chris Volinsky. All together now: A perspective on the Netflix prize. Chance, 23(1):24–29, 2010.

[161] Michel Regenwetter, Jason Dana, and Clintin P Davis-Stober. Transitivity of preferences. Psychological review, 118(1):42, 2011.

[162] Richard D McKelvey. Intransitivities in multidimensional voting models and some implications for agenda control. 1976.

[163] Richard D McKelvey. General conditions for global intransitivities in formal voting models. Econometrica: journal of the Econometric Society, pages 1085–1112, 1979.

[164] Christopher E Zwilling, Daniel R Cavagnaro, Michel Regenwetter, Shiau Hong Lim, Bryanna Fields, and Yixin Zhang. QTest 2.1: Quantitative testing of theories of binary choice using Bayesian inference. Journal of Mathematical Psychology, 91:176–194, 2019.

[165] Kurt Taylor Gaubatz. Intervention and intransitivity: Public opinion, social choice, and the use of military force abroad. World Politics, 47(4):534–554, 1995.

[166] Thomas Flanagan. The staying power of the legislative status quo: Collective choice in canada's parliament after Morgentaler. Canadian Journal of Political Science/Revue canadienne de science politique, 30(1):31–53, 1997.

[167] John C Blydenburgh. The closed rule and the paradox of voting. The Journal of Politics, 33(1):57–71, 1971.

[168] Jennifer Roback Morse. Constitutional rules, political accidents, and the course of history: new light on the annexation of Texas. The Independent Review, 2(2):173–200, 1997.

[169] William H Riker. Liberalism against populism, volume 34. San Francisco: WH Freeman, 1982.

[170] Malthe Munkøe. Cycles and instability in politics. evidence from the 2009 Danish municipal elections. Public Choice, 158(3-4):383–397, 2014.

[171] Michel Regenwetter, Aeri Kim, Arthur Kantor, and Moon-Ho R Ho. The unexpected empirical consensus among consensus methods. Psychological Science, 18(7):629–635, 2007.

[172] Peter Kurrild-Klitgaard. An empirical example of the Condorcet paradox of voting in a large electorate. Public Choice, 107(1-2):135–145, 2001.

[173] William V Gehrlein and Dominique Lepelley. Condorcet efficiency and social homogeneity. In Voting Paradoxes and Group Coherence, pages 157–198. Springer, 2011.

[174] Ralph H Masure and Warder Clyde Allee. The social order in flocks of the common chicken and the pigeon. The Auk, pages 306–327, 1934.

[175] Nicolaus Tideman. Collective decisions and voting: the potential for public choice. Ashgate Publishing, Ltd., 2006.

[176] Adrian Van Deemen. On the empirical relevance of Condorcet's paradox. Public Choice, 158(3-4):311–330, 2014.

[177] Eerik Lagerspetz. Social choice in the real world ii: cyclical preferences and strategic voting in the finnish presidential elections. Scandinavian Political Studies, 20(1):53–67, 1997.

[178] William H Riker, William H Riker, and William H Riker. The art of political manipulation, volume 587. Yale University Press, 1986.

[179] Paul CH Albers and Han de Vries. Elo-rating as a tool in the sequential estimation of dominance strengths. Animal Behaviour, pages 489–495, 2001.

[180] HAN De Vries. Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. Animal Behaviour, 55(4):827–843, 1998.

[181] James Berger. Statistical decision theory: foundations, concepts, and methods. Springer Science & Business Media, 2013.

[182] Jouni Kerman et al. Neutral noninformative and informative conjugate beta and gamma prior distributions. Electronic Journal of Statistics, 5:1450–1470, 2011.

[183] Frank Tuyl, Richard Gerlach, Kerrie Mengersen, et al. Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters. Bayesian analysis, 4(1):151–158, 2009.

[184] John Kruschke. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press, 2014.

[185] Thomas M Cover and Joy A Thomas. Elements of information theory. John Wiley & Sons, 2012.

[186] Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. The annals of mathematical statistics, 9(1):60–62, 1938.

[187] Adolf Buse. The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. The American Statistician, 36(3a):153–157, 1982.

[188] Pragya Sur, Yuxin Chen, and Emmanuel J Candès. The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. Probability Theory and Related Fields, 175(1-2):487–558, 2019.

[189] Hirotugu Akaike. A new look at the statistical model identification. IEEE transactions on automatic control, 19(6):716–723, 1974.

[190] AN Kolmogorov-Smirnov, A Kolmogorov, and M Kolmogorov. Sulla determinazione empírica di uma legge di distribuzione. 1933.

[191] Nickolay Smirnov. Table for estimating the goodness of fit of empirical distributions. The annals of mathematical statistics, 19(2):279–281, 1948.

[192] Frank J Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association, 46(253):68–78, 1951.

[193] Jerzy Neyman and Egon Sharpe Pearson. On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231(694-706):289–337, 1933.

[194] Kenneth J Arrow. Social choice and individual values, volume 12. Yale university press, 2012.

[195] Duncan Black et al. The theory of committees and elections. 1958.

[196] Peter Kurrild-Klitgaard. Voting paradoxes under proportional representation: evidence from eight Danish elections. Scandinavian Political Studies, 31(3):242–267, 2008.

[197] Peter Kurrild-Klitgaard. Trump, Condorcet and Borda: Voting paradoxes in the 2016 Republican presidential primaries. European Journal of Political Economy, 55:29–35, 2018.

[198] Philip J Reny. Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach. Economics Letters, 70(1):99–105, 2001.

[199] Allan Gibbard. Manipulation of voting schemes: a general result. Econometrica: journal of the Econometric Society, pages 587–601, 1973.

[200] Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. Journal of economic theory, 10(2):187–217, 1975.

[201] Gerry Mackie. Democracy defended. Cambridge University Press, 2003.

[202] Ad MA Van Deemen and Noël P Vergunst. Empirical evidence of paradoxes of voting in dutch elections. Public Choice, 97(3):475–490, 1998.

[203] William V Gehrlein. Condorcet's paradox and the Condorcet efficienty of voting rules. Mathematica Japonica, 45:173–199, 1997.

[204] Michel Regenwetter, James Adams, and Bernard Grofman. On the (sample) Condorcet efficiency of majority rule: An alternative view of majority cycles and social homogeneity. Theory and Decision, 53(2):153–186, 2002.

[205] Amartya K Sen. A possibility theorem on majority decisions. Econometrica: Journal of the Econometric Society, pages 491–499, 1966.

[206] Michel Regenwetter, Bernard Grofman, Ilia Tsetlin, and Anthony AJ Marley. Behavioral social choice: probabilistic models, statistical inference, and applications. Cambridge University Press, 2006.

[207] Paul R Abramson, John H Aldrich, Phil Paolino, and David W Rohde. Third-party and independent candidates in American politics: Wallace, Anderson, and Perot. Political Science Quarterly, 110(3):349–367, 1995.

[208] Benjamin Radcliff. Collective preferences in presidential elections. Electoral Studies, 13(1):50–57, 1994.

[209] Courtney Kennedy, Mark Blumenthal, Scott Clement, Joshua D Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, Lee Miringoff, Kristen Olson, Douglas Rivers, et al. An evaluation of the 2016 election polls in the United States. Public Opinion Quarterly, 82(1):1–33, 2018.

[210] Michael C Appleby. Competition in a red deer stag social group: rank, age and relatedness of opponents. Animal Behaviour, 31(3):913–918, 1983.

[211] Filippo Galimberti, Anna Fabiani, and Luigi Boitani. Socio-spatial levels in linearity analysis of dominance hierarchies: a case study on elephant seals. Journal of ethology, 21(2):131–136, 2003.

[212] Keren Klass and Marina Cords. Effect of unknown relationships on linearity, steepness and rank ordering of dominance hierarchies: simulation studies based on data from wild monkeys. Behavioural processes, 88(3):168–176, 2011.

[213] Keren Klass and Marina Cords. Agonism and dominance in female blue monkeys. American journal of primatology, 77(12):1299–1315, 2015.

[214] Laura Muniz, Susan Perry, Joseph H Manson, Hannah Gilkenson, Julie Gros-Louis, and Linda Vigilant. Male dominance and reproductive success in wild white-faced capuchins (Cebus capucinus) at Lomas Barbudal, Costa Rica. American Journal of Primatology, 72(12):1118–1130, 2010.

[215] Mark HJ Nelissen. Structure of the dominance hierarchy and dominance determining "group factors" in Melanochromis Auratus (Pisces, Cichlidae). Behaviour, 94(1-2):85–107, 1985.

[216] Martha M Robbins, Netzin Gerald-Steklis, Andew M Robbins, and H Dieter Steklis. Long-term dominance relationships in female mountain gorillas: strength, stability and determinants of rank. Behaviour, 142(6):779–809, 2005.

[217] Jeff Rushen. The peck orders of chickens: how do they develop and why are they linear? Animal Behaviour, 30(4):1129–1137, 1982.

[218] Carlos Drews. The concept and definition of dominance in animal behaviour. Behaviour, 125(3-4):283–313, 1993.

[219] Andreas Koenig. Competition for resources and its behavioral consequences among female primates. International Journal of Primatology, 23(4):759–783, 2002.

[220] Erling Johan Solberg and Thor Harald Ringsby. Does male badge size signal status in small island populations of house sparrows, passer domesticus? Ethology, 103(3):177–186, 1997.

[221] Hurst Hugh Shoemaker. Social hierarchy in flocks of the canary. The Auk, pages 381–406, 1939.

[222] Lynne A Isbell and Truman P Young. Ecological models of female social relationships in primates: similarities, disparities, and some directions for future clarity. Behaviour, pages 177–202, 2002.

[223] Thorleif Schjelderup-Ebbe. Contributions to the social psychology of the domestic chicken. journal for psychology and physiology of the sensory organs. Dept. 1. journal for psychology, 1922.

[224] HG Landau. On dominance relations and the structure of animal societies: II. some effects of possible social factors. The Bulletin of Mathematical Biophysics, 13(4):245–262, 1951.

[225] Ken Yasukawa and Elyse I Bick. Dominance hierarchies in dark-eyed juncos (junco hyemalis): a test of a game-theory model. Animal Behaviour, 31(2):439–448, 1983.

[226] Linton C Freeman, Sue C Freeman, and A Kimball Romney. The implications of social structure for dominance hierarchies in red deer, Cervus elaphus L. Animal Behaviour, 44:239–245, 1992.

[227] John Maynard Smith and Geoffrey A Parker. The logic of asymmetric contests. Animal behaviour, 24(1):159–175, 1976.

[228] Han de Vries and Michael C Appleby. Finding an appropriate order for a hierarchy: a comparison of the I&SI and the BBS methods. Animal Behaviour, 59(1):239–245, 2000.

[229] Robert Boyd and Joan B Silk. A method for assigning cardinal dominance ranks. Animal Behaviour, 31(1):45–58, 1983.

[230] Han De Vries, Jeroen MG Stevens, and Hilde Vervaecke. Measuring and testing the steepness of dominance hierarchies. Animal Behaviour, 71(3):585–592, 2006.

[231] Patrick Bossuyt. A comparison of probabilistic unfolding theories for paired comparisons data. Springer Science & Business Media, 2012.

[232] WA Thompson and Russell Remage. Rankings from paired comparisons. The Annals of Mathematical Statistics, 35(2):739–747, 1964.

[233] C Alex McMahan and Max D Morris. Application of maximum likelihood paired comparison ranking to estimation of a linear dominance hierarchy in animal societies. Animal behaviour, 32(2):374–378, 1984.

[234] RH Masure and WC Allee. Flock organization of the shell parrakeet Melopsittacus undulatus Shaw. Ecology, 15(4):388–398, 1934.

[235] Mary A Bennett. The social hierarchy in ring doves. Ecology, 20(3):337–357, 1939.

[236] Brandon C Wheeler, Clara J Scarry, and Andreas Koenig. Rates of agonism among female primates: a cross-taxon perspective. Behavioral Ecology, 24(6):1369–1380, 2013.

[237] James Clerk Maxwell. Address to the mathematical and physical section of the British association. 1870.

[238] Alisa Bokulich. Maxwell, Helmholtz, and the unreasonable effectiveness of the method of physical analogy. Studies in History and Philosophy of Science Part A, 50:28–37, 2015.

[239] Hong Qian. Vector field formalism and analysis for a class of thermal ratchets. Physical review letters, 81(15):3063, 1998.

[240] Nicolaas Godfried Van Kampen. Stochastic processes in physics and chemistry, volume 1. Elsevier, 1992.

[241] David F Anderson and Thomas G Kurtz. Stochastic analysis of biochemical systems, volume 1. Springer, 2015.

[242] Daniel T Gillespie. Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem., 58:35–55, 2007.

[243] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. The journal of physical chemistry A, 104(9):1876–1889, 2000.

[244] Patrick Billingsley. Probability and measure. John Wiley & Sons, 2008.

[245] Jagpreet Chhatwal, Suren Jayasuriya, and Elamin H Elbasha. Changing cycle lengths in state-transition models: challenges and solutions. Medical Decision Making, 36(8):952–964, 2016.

[246] S Unnikrishna Pillai, Torsten Suel, and Seunghun Cha. The Perron-Frobenius theorem: some of its applications. IEEE Signal Processing Magazine, 22(2):62–75, 2005.

[247] Semyon Aranovich Gershgorin. Uber die abgrenzung der eigenwerte einer matrix. Proceedings of the Russian Academy of Sciences. Mathematical Series, (6):749–754, 1931.

[248] Martin J Klein. Principle of detailed balance. Physical Review, 97(6):1446, 1955.

[249] Ashok K Chandra, Prabhakar Raghavan, Walter L Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times. Computational Complexity, 6(4):312–340, 1996.

[250] Brad H McRae, Brett G Dickson, Timothy H Keitt, and Viral B Shah. Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology, 89(10):2712–2724, 2008.

[251] László Lovász et al. Random walks on graphs: A survey. Combinatorics, Paul Erdos is eighty, 2(1):1–46, 1993.

[252] Crawford S Holling. Resilience and stability of ecological systems. Annual review of ecology and systematics, 4(1):1–23, 1973.

[253] Massimiliano Esposito, Katja Lindenberg, and Christian Van den Broeck. Universality of efficiency at maximum power. Physical review letters, 102(13):130602, 2009.

[254] Hong Qian. Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. Physical Review E, 63(4):042103, 2001.

[255] Thomas M Cover and J Halliwell. Which processes satisfy the second law. Physical origins of time asymmetry, pages 98–107, 1994.

[256] Alfréd Rényi et al. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California, 1961.

[257] Da-Quan Jiang, Donghua Jiang, and Min Qian. Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems. Number 1833. Springer Science & Business Media, 2004.

[258] LS Garcia-Colin and FJ Uribe. Extended irreversible thermodynamics beyond the linear regime. a critical overview. J. Non-Equilib. Thermodyn, 16(2):89–128, 1991.

[259] Maurice Antony Biot et al. Linear thermodynamics and the mechanics of solids. In Proceedings of the Third US National Congress of Applied Mechanics, American Society of Mechanical Engineers. Citeseer, 1958.

[260] Lars Onsager. Reciprocal relations in irreversible processes. I. Physical review, 37(4):405, 1931.

[261] Lars Onsager. Reciprocal relations in irreversible processes. II. Physical review, 38(12):2265, 1931.

[262] Shumpei Yamamoto, Sosuke Ito, Naoto Shiraishi, and Takahiro Sagawa. Linear irreversible thermodynamics and onsager reciprocity for information-driven engines. Physical Review E, 94(5):052121, 2016.

[263] Roger A Horn, Roger A Horn, and Charles R Johnson. Topics in matrix analysis. Cambridge university press, 1994.

[264] Hakop Hakopian, Kurt Jetter, and Georg Zimmermann. Vandermonde matrices for intersection points of curves. Jaen J. Approx, 1(1):67–81, 2009.

[265] Edward K Agarwala, Hillel J Chiel, and Peter J Thomas. Pursuit of food versus pursuit of information in a markovian perception–action loop model of foraging. Journal of theoretical biology, 304:235–272, 2012.

[266] Nihat Ay and Thomas Wennekers. Dynamical properties of strongly interacting markov chains. Neural Networks, 16(10):1483–1497, 2003.

[267] Thomas Wennekers and Nihat Ay. Finite state automata resulting from temporal information maximization and a temporal learning rule. Neural computation, 17(10):2258–2290, 2005.

[268] Christopher V Rao and Adam P Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. The Journal of chemical physics, 118(11):4999–5010, 2003.

[269] Nicolaas G Van Kampen. Itô versus Stratonovich. Journal of Statistical Physics, 24(1):175–187, 1981.

[270] John Smythe, Frank Moss, Peter VE McClintock, and Douglas Clarkson. Ito versus Stratonovich revisited. Physics Letters A, 97(3):95–98, 1983.

[271] Yousef Alnafisah. The implementation of Milstein scheme in two-dimensional SDEs using the Fourier method. In Abstract and Applied Analysis, volume 2018. Hindawi, 2018.

[272] Hye-Won Kang, Likun Zheng, and Hans G Othmer. A new method for choosing the computational cell in stochastic reaction–diffusion systems. Journal of mathematical biology, 65(6-7):1017–1099, 2012.

[273] Samuel A Isaacson. A convergent reaction-diffusion master equation. The Journal of chemical physics, 139(5):054101, 2013.

[274] Qian Minping and Qian Min. Circulation for recurrent markov chains. Journal of Probability Theory and Related Areas, 59(2):203–210, 1982.

[275] Eitan Y Levine and Baruch Meerson. Impact of colored environmental noise on the extinction of a long-lived stochastic population: Role of the allee effect. Physical Review E, 87(3):032127, 2013.

[276] Alexander G Strang, Karen C Abbott, and Peter J Thomas. How to avoid an extinction time paradox. Theoretical Ecology, 12(4):467–487, 2019.

[277] Joseph Xu Zhou, MDS Aliyu, Erik Aurell, and Sui Huang. Quasi-potential landscape in complex multi-stable systems. Journal of the Royal Society Interface, 9(77):3539–3553, 2012.

[278] Otso Ovaskainen and Baruch Meerson. Stochastic models of population extinction. Trends in ecology & evolution, 25(11):643–652, 2010.

[279] Carl M Bender and Steven A Orszag. Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory. Springer Science & Business Media, 2013.

[280] James A Sethian and Alexander Vladimirsky. Ordered upwind methods for static hamilton–jacobi equations. Proceedings of the National Academy of Sciences, 98(20):11069–11074, 2001.

[281] Kazuhisa Tomita and Hiroyuki Tomita. Irreversible circulation of fluctuation. Progress of Theoretical Physics, 51(6):1731–1749, 1974.

[282] Juan Pablo Gonzalez, John C Neu, and Stephen W Teitsworth. Experimental metrics for detection of detailed balance violation. Physical Review E, 99(2):022143, 2019.

[283] Tyler Lu and Craig Boutilier. Budgeted social choice: From consensus to personalized decision making. In Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

[284] Jean-François Laslier and Karine Van der Straeten. A live experiment on approval voting. Experimental Economics, 11(1):97–105, 2008.

[285] Guy Sella and Aaron E Hirsh. The application of statistical physics to evolutionary biology. Proceedings of the National Academy of Sciences, 102(27):9541–9546, 2005.

[286] NH Barton and JB Coe. On the application of statistical physics to evolutionary biology. Journal of theoretical biology, 259(2):317–324, 2009.

[287] Harold P De Vladar and Nicholas H Barton. The contribution of statistical physics to evolutionary biology. Trends in ecology & evolution, 26(8):424–432, 2011.

[288] Ville Mustonen and Michael Lässig. Fitness flux and ubiquity of adaptive evolution. Proceedings of the National Academy of Sciences, 107(9):4248–4253, 2010.