Conformal geometry of Markov chains

Peter G. Doyle Gregory Leibon Jean Steiner

Version 4A1 dated 4 November 2009
GNU FDL*

Abstract

A conformal deformation of a Markov chain changes its equilibrium measure without changing the equilibrium transition rates between states. The expected time to traverse an oriented cycle of states is a conformal invariant, and taken together these cycle times determine the conformal equivalence class of the chain. The same information is provided by the matrix of commuting times between pairs of states, together with the reversibility obstruction, which measures the excess time to traverse a cycle in one direction rather than the other. The states of any Markov chain can be embedded in a Euclidean so that the squared distance between states is the commute time. We find a minimax characterization of commute times, and from this we get monotonicity of commute times with respect to equilibrium transition rates.

1 Conformal equivalence of Markov chains

Let ${P_i}^j$ be the transition matrix of an ergodic discrete-time Markov chain on states $\{1, \ldots, r\}$. Let w^i with $\sum_i w_i = 1$ be its equilibrium measure, so that

$$\sum_i w_i {P_i}^j = w^j.$$

The upper index of w^i indicates that we are thinking of w as a row vector, with

$$wP = w.$$

Define the *charge operator*

$$\Delta^{ij} = w^j(I_i{}^j - P_i{}^j).$$

The upper indices of Δ^{ij} indicate that we are thinking of Δ as defining a bilinear form $\Delta(x, y)$, which we'll call the *charge form*, where for column vectors x_i, y_j

$$\Delta(x, y) = \sum_{i,j} x_i \Delta^{ij} y_y.$$

Because

$$\sum_i \Delta^{ij} = \sum_j lap^{ij} = 0$$

this bilinear form pushes down to the quotient $\mathbf{R}^n/1$ where we mod out by constant vectors. This is where the charge form wants to live.

This charge operator a close relative of the *Laplacian* $I - P$. The distinction is that the Laplacian maps functions to functions (column vectors to column vectors), while the charge operator maps functions to measures (column vectors to row vectors).

We say that two Markov chains are *comformally equivalent* if they have the same charge operator, which is the same as having the same charge form. Probabilistically, two chains are conformally equivalent just if for every pair of states $i \neq j$ they have the same equilibrium transition rate $w_i P_i{}^j$. This definition is meant to correspond with the notion of conformal equivalence for Riemannian surfaces, with the difference that here we impose what amounts to a 'volume condition', whereby conformally equivalent surfaces would be required to have the same total area.

To get a Markov chain $\bar{P}_i{}^j$ conformally equivalent to $P_i{}^j$ we choose constants c_i normalized so that $\sum_i c_i w^i = 1$ and set

$$\bar{P}_i{}^j = I_i{}^j - \frac{1}{c_i}(I_i{}^j - P_i{}^j),$$

so that

$$I_i{}^j - \bar{P}_i{}^j = \frac{1}{c_i}(I_i{}^j - P_i{}^j).$$

Then the equilibrium measure $\bar{w}^i$ for $\bar{P}_i{}^j$ will be $c_i w^i$ because

$$\sum_i c_i w^i (I_i{}^j - \bar{P}_i{}^j) = \sum_i w^i (I_i{}^j - P_i{}^j) = 0.$$

Thus

$$\bar{\Delta}^{ij} = \bar{w}^i (I_i{}^j - \bar{P}_i{}^j) = c_i w^i \frac{1}{c_i} (I_i{}^j - P_i{}^j) = \Delta^{ij},$$

so $\bar{P}$ and P are conformally equivalent. The only proviso here is that we must take care that the diagonal entries $\bar{P}_i{}^i$ stay positive. This is a headache associated with working in discrete time. To avoid it, we could work with continuous-time chains. Or we could define two chains to be conformally equivalent if their charge operators agree up to a constant multiple. This would correspond well with the usual notion of conformal equivalence in geometry. But then we'd have to keep using the phrase 'up to a constant multiple', so let's just ignore this issue.

2 Overview

The *commute time* T_{ab} between two states a, b of an ergodic Markov chain is the expected time, starting from a, to go to b and then back to a. As we will see, commute time are conformal invariants. More generally, the expected time to traverse any oriented cycle of states is a conformal invariant. Taken together these cycle times form a complete set of conformal invariants. In fact, we need only consider cycles of length 2 and 3 If M_{ij} is the expected time to get from i to j, then knowing the commute times

$$T_{ab} = M_{ab} + M_{ba}$$

and the triangle times

$$T_{abc} = M_{ab} + M_{bc} + M_{ca}$$

will determine the conformal class of the chain. In fact, in place of the triangle times T_{abc} it will suffice to know the triangle difference $D_{abc} = T_{abc} - T_{acb}$, because from D_{abc} together with the commute times T_{ab}, T_{bc}, T_{ac} we can recover T_{abc}:

$$T_{abc} = \frac{1}{2}(D_{abc} + T_{ab} + T_{bc} + T_{ca}).$$

The triangle differences constitute a cohomology class which we'll call the *reversibility obstruction*. This class vanishes just if the chain is time-reversible. So the upshot is that the commute times and the reversibility obstruction together capture the conformal class of a chain.

To portray the commute times, we will show how to embed the states of a chain with r states in Euclidean $r-1$-space so that the squared distance between states is the commute time. Elsewhere [**?**] we have given a stripped-down description of this. Here we take a thoroughgoing, conformally correct approach. As a payoff, we get a natural minimax characterization of commute times. This leads to the monotonicity law for commute times: If all equilibrium interstate transition rates are increased, then all commute times are diminished. For time-reversible chains, this monotonicity law is an ancient and powerful tool. It is questionable how useful it will prove to be in the general case.

3 Visualizing commute times

One way to determine the expected commute time T_{ab} between a and b is to run the chain for a long time T (beware of confusion!), paying attention to when the chain is at a or b and ignoring other states. If R is the number of runs of a's (which is within 1 of the number of runs of b's), then

$$T_{ab} \approx T/R.$$

To keep track of R we imagine painting our Markovian particle green when it reaches a and red when it reaches b. Let r_{ab} be the equilibrium rate at which red particles are being painted green. Ignoring end effects, over our long time interval T, R above is the number of times a red particle gets painted green (and vice versa), and thus roughly Tr_{ab}, and it follows that

$$T_{ab} = \frac{1}{r_{ab}}.$$

This is an instance of the general principle from renewal theory that when events happen at rate r, the expected time between events is $1/r$.

Note. This painting business is very close to a model developed by Kingman [8] and Kelly [7]. (See exercise 1 in section 3.3 of Doyle and Snell [5].) However, I don't know that Kingman and Kelley ever made the connection

to commute times, and it is possible that their discussion concerned only time-reversible chains. Somebody should check this.

It is high time to observe that if $\hat{T}_{ab}$ is the commute time for the time-reversed chain (according to the general convention that time-reversed quantities wear hats), we have

$$T_{ab} = T_{ba} = \hat{T}_{ab} = \hat{T}_{ba}.$$

We claim to be able to see this from our way of approximating T_{ab} by observing the chain over a long time. If we reverse a record of the chain moving forward for a long time, we see roughly a record of the time-reversed chain starting in equlibrium. In fact if we started the original chain in equilibrium we're golden. If we started the chain not in equilibirum (e.g. by starting at a, as we might well be tempted to do), there will be problems toward the end of the time-reversed record, as the time-reversed chain gets drawn to end where the forward chain began. But this effect is negligible when T is large.

4 The Laplacian and the cross-potential

Consider a discrete-time Markov chain with transition probabilities

$$P_i{}^j = \text{Prob(at } j \text{ next time|at } i \text{ this time)}.$$

Assume the chain is ergodic, so that there is a unique equilibrium measure w^i with

$$\sum_i w^i P_i{}^j = w^j,$$

$$\sum_i w^i = 1.$$

Define the *Laplacian*

$$D^{ij} = w^i(I_i{}^j - P_i{}^j).$$

For $i \neq j$, $-D^{ij}$ tells the equilibrium rate of transitions from i to j; D^{ii} tells the total rate of transitions to and from states other than i.

Note. Really $I - P$ is the Laplacian, and that this is not the Laplacian, but the matrix of the Dirichlet form. But we're going to call this the Laplacian anyway.

The time-reversed Markov chain has Laplacian $\hat{D}^{ij} = D^{ji}$. A time-reversible chain has $D^{ij} = D^{ji}$.

We have

$$\sum_i D^{ij} = \sum_j D^{ij} = 0.$$

So considered as a matrix, D^{ij} is not invertible. However, it has a generalized inverse Z_{ij} with the property that for any measure of total mass 0, which is to say for any u^i with $\sum_i u^i = 0$, we have

$$\sum_{jk} u^j Z_{jk} D^{kl} = u^l$$

and

$$\sum_{jk} D^{ij} Z_{jk} u^k = u^i.$$

An equivalent way to write this is

$$\sum_{jk} D^{ij} Z_{jk} D^{kl} = D^{il},$$

because if we think of D^{ij} as a matrix, its rows and columns both span the space of measures with total mass 0.

A sensible choice for the generalized inverse Z_{ij} is

$$Z_{ij} = \frac{1}{w^j} Z_i{}^j$$

where

$$Z_i{}^j = (I_i{}^j - w^j) + (P_i{}^j - w^j) + ((P^2)_i{}^j - w^j) + \ldots,$$

where $(P^2)_i{}^j = \sum_k P_i{}^k P_k{}^j$ represents the matrix square of $P_i{}^j$, and the elided terms involve higher matrix powers. Define $(P^\infty)_i{}^j = w^j$, to suggest that the 'infinitieth power' of $P_i{}^j$ has all rows equal to the vector w^i. We can write

$$\begin{aligned} Z &= (I - P^\infty) + (P - P^\infty) + (P^{(2)} - P^\infty) + \ldots \\ &= (I - P + P^\infty)^{-1} - P^\infty. \end{aligned}$$

This naturally translates into the formula we've given for $Z_i{}^j$, and from there, by 'lowering the index j', we get Z_{ij}.

For this choice of Z we have the natural interpretation that $Z_i{}^j$ is the expected excess number of visits to j for a chain starting at i compared to a chain starting in equilibrium. For the time-reversed chain we get

$$\hat{Z}_{ij} = Z_{ji},$$

and so in particular if the chain is time-reversible we have $Z_{ij} = Z_{ji}$.

For this particular choice of Z_{ij} we have

$$\sum_{jk} Z_{ij} D^{jk} Z_{kl} = Z_{il}.$$

This is all very well, but we still do not want to prescribe this particular choice of Z because it is not conformally invariant: It depends on the equilibrium measure w^i, and not just on the 'matrix' D^{ij}. This makes it insufficiently canonical for us.

What *is* canonical is the bilinear form

$$B(u, v) = \sum_{ij} u^i Z_{ij} v^j$$

when u and v are restricted to the subspace S of measures of total mass 0:

$$S = \{u^i : \sum_i u^i = 0\}$$

Note that the form B does not change if we replace Z_{ij} with $Z_{ij} + \alpha_i + \beta_j$.

Fixing a, b, c, d and setting

$$u = \delta_a{}^i - \delta_b{}^i;\ v = \delta_c{}^i - \delta_d{}^i$$

gives us the *cross-potential*

$$N_{abcd} = B(\delta_a{}^i - \delta_b{}^i, \delta_c{}^i - \delta_d{}^i) = Z_{ac} - Z_{ad} - Z_{bc} + Z_{bd}.$$

N satisfies

$$N_{bacd} = N_{abdc} = -N_{abcd}.$$

For the time-reversed process

$$\hat{N}_{abcd} = N_{cdab}.$$

Clearly, knowing N is the same as knowing B, or D. If we know w as well as N we can recover our sensible-but-not-canonical Z:

$$Z_{ij} = \sum_{kl} N_{ikjl} w^k w^l.$$

Different choices of w in this formula lead to different Z's, but they all determine the same bilinear form B. From Z and w we can recover P.

In general, it is useful to think of an ergodic Markov chain as specified by the cross-potential N, which determines its conformally invariant properties, together with the equilibrium measure w. Expressing formulas in these terms allows us to see the extent to which quantities are conformally invariant (like N, B, and D) or not (like w, Z, P).

Complaint. N and w together don't quite determine the original transition rates for a continuous-time Markov chain, or rather, they wouldn't do so if we had some way to distinguish between remaining at i and moving from i to i. Such a distinction is not possible for discrete-time chains represented by matrices, but we could handle it in the continuous case by allowing for non-zero transition rates on the diagonal. Better yet, we could reformulate Markov chain theory in the context of queuing networks based on 1-complexes (graphs where loops and multiple edges are allowed). This would give us a way to distinguish different ways of stepping from i to j. A further step would be to allow a general distribution for the time it takes to make a transition for i to j. This would be very helpful when watching the chain only when it is in a subset of its states, as in the case above where we contemplated watching the chain only when it is at a or b. We didn't say just what we meant by this, because it doesn't conveniently fit into the usual formulation of Markov chain theory.

5 Probabilistic and electrical interpretation

We may interpret N_{abcd} probabilistically as the equilibrium concentration difference between c and d due to a unit flow of particles entering at a and leaving at b. Here's what this means. Introduce Markovian particles at a at a unit rate, and remove them when they reach b. Write the 'dynamic equilibrium' measure of particles at i as $w^i \phi_i$, so that ϕ_i tells the concentration of particles relative to the 'static equilibrium' measure w^i. Conservation of

particles implies that

$$w^i \phi_i \sum_j P_i{}^j - \sum_j w^j \phi_j P_j{}^i = \delta_a{}^i - \delta_b{}^i.$$

We hasten to rewrite this in the conformally correct form

$$\sum_j \phi_j D^{ji} = \delta_a{}^i - \delta_b{}^i.$$

Since also

$$\sum_j (Z_{aj} - Z_{bj}) D^{ji} = \delta_a{}^i - \delta_b{}^i$$

and since D kills only constants, if follows that

$$\phi_j = Z_{aj} - Z_{bj} + C,$$

and thus

$$\phi_c - \phi_d = Z_{ac} - Z_{bc} - Z_{ad} + Z_{bd} = N_{abcd}.$$

From this probabilistic interpretation of N we can see that $N_{abab} = T_{ab}$, the commute time between a and b. Indeed, in the particle-painting scenario introduced earlier, T_{ab} is the reciprocal of the rate at which red particles are turning green at a. Paying attention only to green particles, we see green particles appearing at a at rate $1/T_{ab}$, and disappearing at b. The equilibrium concentration of green particles at i is the probability p_i of hitting a before b for the time-reversed chain, and in particular $p_a = 1$ and $p_b = 0$, so the concentration difference between a and b is 1. Multiplying the green flow by T_{ab} normalizes it to a unit flow with concentration difference T_{ab} between a and b. So

$$T_{ab} = N_{abab}.$$

If we embellish this probabilistic scenario by imagining that our particles carry a positive charge, we may identify the net flow of particles with electrical current; the concentration of particles (relative to the equilibrium measure) with electrical potential; and differences of concentration with voltage drop. With this terminology, N_{abcd} tells the voltage drop between c and d due to a unit current from a to b. Traditionally this way of talking is reserved for time-reversible Markov chains, which are precisely those for which we have the 'reciprocity law' $N_{abcd} = N_{cdab}$. For such chains, if we build a resistor network where nodes $i \neq j$ are joined by a resistor of conductance

(i.e., reciprocal resistance) $-D^{ij}$, then N_{abcd} will indeed be the voltage drop between c and d due to a unit current from a to b. We propose to extend this way of talking to non-time-reversible chains.

In electrical terms, the voltage drop N_{abab} between a and b due to a unit current between a and b is the *effective resistance*. This is the same as the reciprocal of the current that flows when a 1-volt battery is connected up between a and b—which is what we get in effect when we measure commute times using green and red paint. So the commute time $T_{ab} = N_{abab}$ is the same as the effective resistance between a and b.

The connection of commute time to effective resistance, and the general recognition that commute times play a key role in understanding Markov chains, is due to Chandra et al. [1].

Note. Now we are in a position to understand the significance of the name 'cross-potential'. This name is meant to indicate the connection of N_{abcd} to the cross-ratio of complex function theory. If we extend our notions about Markov chains to cover Brownian motion on the Riemann sphere, we get

$$\begin{aligned} N_{abcd} &= -\frac{1}{2\pi}(\log|a-c| - \log|a-d| - \log|b-c| + \log|b-d|) \\ &= -\frac{1}{2\pi}\log\left|\frac{a-c}{a-d}\frac{b-d}{b-c}\right|) \\ &= -\frac{1}{2\pi}\Re\log\frac{a-c}{a-d}\frac{b-d}{b-c}. \end{aligned}$$

We don't have to specify a metric on the sphere here, because Laplace's equation is a conformal invariant in two dimensions. Thinking of the sphere as being an electrical conductor with constant conductivity (say, 1 mho 'per square'), the electrical interpretation becomes exact. The advantage of having N to take four 'arguments' now becomes apparent, because $N_{abcb} = \infty$. That's why engineers using look for cracks in nuclear reactor cooling pipes with a *4-point probe*. To get a sensible generalization of T_{ab} we will need to do some kind of renormalization, which will introduce a dependence on the metric. We should not be sorry about this, because it brings curvature into the picture—and you know that can't be bad.

6 Realization

Now, finally, to realize commute times as squared distances. From the bilinear form B we get the quadratic form

$$Q(u) = ||u||^2 = B(u,u) = \sum_{ij} u^i Z_{ij} u^j.$$

$$T_{ab} = N_{abab} = Q(\delta_a{}^i - \delta_b) = ||\delta_a - \delta_b||^2.$$

So if we map i to δ_i then the commute time T_{ab} becomes the squared distance between the images in the Q-norm.

That is, if what we're calling the Q-norm is indeed a norm. Is Q really positive definite?

To understand better what is going on here, it is useful to look at the bilinear form

$$L(\phi,\psi) = \sum_{ij} \phi_i D^{ij} \psi_j,$$

where we think of ϕ and ψ as being defined only modulo additive constants. If we think of ϕ_i as the potential of the measure

$$\sum_i \phi_i D^{ik},$$

then this is the same bilinear form as before, except that now instead of measures of total mass 0 it takes as its arguments the corresponding potentials, the first with respect to the original chain, and the second with respect to the time-reversed chain:

$$L(\phi,\psi) = B(\sum_i \phi_i D^{ik}, \sum_i \psi_i D^{ki}) = B(\sum_i \phi_i D^{ik}, \sum_i \hat{D}^{ik} \psi_i).$$

This follows from the formula $DZD = D$ above.

Now to get the equivalent of Q in this context we restrict to the subspace

$$V = \{(\phi,\psi) : \sum_i \phi_i D^{ik} = \sum_j D^{kj} \psi_j\}$$

and take as our quadratic form

$$R((\phi,\psi)) = L(\phi,\psi).$$

In the case of a time-reversible chain, V is just the diagonal $\phi = \psi$, and

$$Q(\phi D) = R((\phi, \phi)) = L(\phi, \phi) = \sum_{ij} \phi_i D^{ij} \phi_j = \frac{1}{2} \sum_{ij} (-D^{ij})(\phi_i - \phi_j)^2.$$

This is evidently positive-definite. Indeed, if we associate to (ϕ, ϕ) the vector with $\binom{r}{2}$ coordinates $\sqrt{-D^{ij}}(\phi_i - \phi_j)$, $i < j$, then we will have embedded the normed space (V, R), and along with it our Markov chain, in Euclidean $\binom{r}{2}$-space.

Electrically, what we have done here is to account for the energy being dissipated in the network by adding up the energy dissipated by individual resistors. And there should be some kind of probabilistic interpretation as well.

That's how it works for time-reversible chains, for which $D^{ij} = D^{ji}$. However, the argument extends to the general case by what amounts to a trick. The key is the observation that for $(\phi, \psi) \in V$ we have

$$L(\phi, \psi) = L(\phi, \phi) = L(\psi, \psi).$$

(But please note that in general $L(\phi, \psi) \neq L(\psi, \phi)$!) So

$$Q(\phi D) = R((\phi, \psi)) = L(\phi, \psi) = L(\phi, \phi) = \sum_{ij} \phi_i D^{ij} \phi_j = \frac{1}{2} \sum_{ij} (-D^{ij})(\phi_i - \phi_j)^2.$$

So there is the positive-definiteness we need.

Now, though, we don't see any natural way to interpret the terms of the sum electrically or probabilistically. (Which is not to say that there isn't one!) In putting ϕ in both slots of L we leave the subspace V, and thereby commit what appears to be an unnatural act. But it seems to have paid off.

7 What just happened

We want to explain the proof we have just given in more conceptual terms.

Let V be a finite-dimensional real vector space, and $V^\star$ the dual space, consisting of linear functionals $\phi : V \to \mathbf{R}$. For $u \in V^\star$, $x \in V$ write

$$\langle u, x \rangle_V = u(x)$$

for the natural pairing between V and $V^\star$. Identify V with $V^{\star\star}$ as usual:

$$\langle x, u \rangle_{V^\star} = u(x) = \langle u, x \rangle_V.$$

To a map $f : V \to W$ we associate the adjoint map $f^\star : W^\star \to V^\star$, such that for $u \in W^\star$, $x \in V$

$$\langle f^\star(u), x\rangle_V = u(f(x)).$$

A bilinear form on V arises from a linear map

$$\phi : V \to V^\star$$

via

$$L_\phi(x, y) = \langle \phi(x), y\rangle_V.$$

The adjoint map

$$\phi^\star : V^\star \to V$$

yields the transposed bilinear form

$$L_{\phi^\star}(x, y) = \langle \phi^\star(x), y\rangle_V = \langle x, \phi(y)\rangle_{V^\star} = \langle \phi(y), x\rangle_V = L_\phi(y, x).$$

If ϕ is invertible the inverse

$$\phi^{-1} : V^\star \to V$$

yields the form $L_{\phi^{-1}}$ on $V^\star$:

$$L_{\phi^{-1}}(u, v) = \langle \phi^{-1}(u), v\rangle_{V^\star} = \langle v, \phi^{-1}(u)\rangle_V.$$

The forms $L_{\phi^\star}$ and $L_{\phi^{-1}}$ are conjugate, because

$$L_{\phi^{-1}}(u, v) = \langle v, \phi^{-1}(u)\rangle_V = L_\phi(\phi^{-1}(v), \phi^{-1}(u)) = L_{\phi^\star}(\phi^{-1}(u), \phi^{-1}(v)).$$

Going back the other way,

$$L_{\phi^\star}(x, y) = L_{\phi^{-1}}(\phi(x), \phi(y)).$$

From these two equations, we get two distinct ways to conjugate L_ϕ to $L_{\phi^{-1\star}}$. Plugging $\phi = (\phi^{-1})^{-1}$ into the first and putting (x, y) for (u, v), we get

$$L_\phi(x, y) = L_{\phi^{-1\star}}(\phi(x), \phi(y)).$$

Plugging $\phi = (\phi^\star)^\star$ into the second we get

$$L_\phi(x, y) = L_{\phi^{-1\star}}(\phi^\star(x), \phi^\star(y)).$$

Now putting $\phi^\star$ for ϕ we see that in fact there were two ways to conjugate $L_{\phi^{-1}}$ to $L_{\phi^\star}$:

$$L_{\phi^\star}(x,y) = L_{\phi^{-1}}(\phi(x),\phi(y)) = L_{\phi^{-1}}(\phi^\star(x),\phi^\star(y)).$$

Having two ways to conjugate L_ϕ to $L_{\phi^{-1\star}}$ gives us an automorphism $\phi^{-1} \circ \phi^\star$ of L_ϕ:

$$L_\phi(x,y) = L_\phi(\phi^{-1}(\phi^\star(x)),\phi^{-1}(\phi^\star(y))).$$

Along with $\phi^{-1} \circ \phi^\star$ we also have the inverse automorphism $\phi^{-1\star} \circ \phi$:

$$L_\phi(x,y) = L_\phi(\phi^{-1\star}(\phi(x)),\phi^{-1\star}(\phi(y))).$$

We could also consider powers other than -1 of our automorphism, but we don't need to, because the conjugacy between L_ϕ and $L_{\phi^\star}$ is canonical (in the sense of being equivariant with respect to taking duals and inverses) up to this factor of two. The difference between them, as measured by the automorphism $\phi^{-1} \circ \phi^\star$, measures the antisymmetry of L_ϕ. It is destined to play an important role in our future.

Looking now at the level of quadratic forms $Q_\phi(x) = L_\phi(x,x)$, everything in sight is conjugate:

$$Q_\phi(x) = Q_{\phi^\star}(x);$$
$$Q_{\phi^{-1}}(u) = Q_{\phi^{-1\star}}(u) = Q_\phi(\phi^{-1}(u)) = Q_\phi(\phi^{-1\star}(u)).$$

All this nonsense can be made much more concrete using matrices. Let $V = \mathbf{R}^n$ and represent $x \in V$, $u \in V^\star$ as column and row vectors respectively, so that the pairing is just multplying a row vector by a column vector:

$$\langle u, x \rangle_V = ux.$$

Denote transposition of matrices by $\star$. Write

$$L_\phi(x,y) = x^\star A y,$$

so that

$$\phi(x) = x^\star A = (A^\star x)^\star.$$

Now

$$\phi^{-1}(u) = (uA^{-1})^\star = A^{-1\star}u^\star,$$

so

$$L_{\phi^{-1}}(u,v) = \langle v, \phi^{-1}(u)\rangle_V = vA^{-1\star}u^\star = uA^{-1}v^\star.$$

Good!

Now to see the two conjugacies of $L_{\phi^\star}$ with $L_{\phi^{-1}}$:

$$A^\star A^{-1} A = A^\star;$$

$$AA^{-1}A^\star = A^\star.$$

These combine to give two automorphisms of L_ϕ:

$$(A^{-1}A^\star)^\star A(A^{-1}A^\star) = AA^{-1\star}AA^{-1}A^\star = A;$$

$$(A^{-1\star}A)^\star A(A^{-1\star}A) = A^\star A^{-1}A^{-1\star}A = A.$$

Hmm. Why didn't we do it this way in the first place?

So, here's what happened with our Markov chain. We started with the space $V = \mathbf{R}^n/1$ with quadratic form $L_\phi(x,y) = \sum_{ij} x_i D^{ij} y_j$, embedded the states in $V^\star = \mathbf{R}^n \perp 1$ with quadratic form $L_{\phi^{-1}}(u,v) = \sum_{ij} u^i Z_{ij} v^j$, and proved that $L_{\phi^{-1}}$ is positive definite by showing that it is conjugate to L_ϕ.

8 Minimax characterization of commute times and hitting probabilities

Fix states $a \neq b$, and let

$$S_{a,b} = \{\phi | \phi_a = 1, \phi_b = 0\}$$

Here we really should be thinking of ϕ as being defined only up to an additive constant, which means we should write $\phi_a - \phi_b = 1$, but we're going to be sloppy about this, because we want to focus attention on two distinguished elements of $S_{a,b}$ which are naturally 1 and a and 0 at b. These are

$$\bar{\phi}_i = \text{Prob(hit } a \text{ before } b \text{ starting at } i \text{ going backward in time)}$$

and

$$\bar{\psi}_i = \text{Prob(hit } a \text{ before } b \text{ starting at } i \text{ going forward in time).}$$

We've met $\bar{\phi}$ before: It's proportional to the equilibrium concentration of green particles in our painting scenario. $\bar{\psi}$ is the analogous quantity for the reversed chain. The pair $(\bar{\phi}, \bar{\psi})$ belongs to our subset V, because

$$(\bar{\phi}D)^i = (D\bar{\psi})^i = r_{ab}(\delta_a{}^i - \delta_b{}^i).$$

Here we once again are writing $r_{ab} = \frac{1}{T_{ab}}$ for the equilibrium rate of commute between a and b. Observe that any f we have

$$L(\bar{\phi}, f) = L(f, \bar{\psi}) = r_{ab}(f_a - f_b).$$

So whenever f is in $S_{a,b}$ we have

$$L(\bar{\phi}, f) = L(f, \bar{\psi}) = r_{ab},$$

and in particular

$$L(\bar{\phi}, \bar{\psi}) = r_{ab}.$$

Theorem 1

$$r_{ab} = \frac{1}{T_{ab}} \min_{\alpha} \max_{\phi+\psi=2\alpha} L(\phi, \psi).$$

Here and below, α, ϕ, and ψ are restricted to lie in $S_{a,b}$, i.e. to take value 1 at a and 0 at b.

Proof. Whatever α is, we may take $\phi = \bar{\phi}$ (and thus $\psi = 2\alpha - \bar{\phi}$), and have

$$L(\phi, \psi) = L(\bar{\phi}, \psi) = r_{ab}$$

as above. So

$$\min_{\alpha} \max_{\phi+\psi=2\alpha} L(\phi, \psi) \geq r_{ab}.$$

To prove the inequality in the other direction, and in the process identify where the minimax is achieved, take

$$\alpha = \frac{1}{2}(\bar{\phi} + \bar{\psi}).$$

If $\phi + \psi = 2\alpha$ then we can write

$$\phi = \bar{\phi} + f$$

and

$$\psi = \bar{\psi} - f,$$

where $f_a = f_b = 0$.

Now

$$L(\bar{\phi}, f) = L(f, \bar{\psi}) = r_{ab}(f_a - f_b) = 0,$$

so

$$L(\phi, \psi) = L(\bar{\phi} + f, \bar{\psi} - f) = L(\bar{\phi}, \bar{\psi}) - L(f, f) = r_{ab} - L(f, f).$$

And even though we claim it is a travesty to put the same f into both slots of L, we still have

$$L(f, f) \geq 0 :$$

That was the upshot of our embedding investigation. So

$$L(\phi, \psi) \leq r_{ab},$$

still assuming $\alpha = \frac{1}{2}(\bar{\phi} + \bar{\psi})$ and $\phi + \psi = 2\alpha$. Hence

$$\min_{\alpha} \max_{\phi+\psi=2\alpha} L(\phi, \psi) \geq r_{ab}. \quad \blacksquare$$

In the time-reversible case, where $D^{ij} = D^{ji}$, this minimax can be reduced to a straight minimum. That's because in this case for any g, f we have $L(f, g) = L(g, f)$, and hence

$$L(g + f, g - f) = L(g, g) - L(f, f).$$

So to maximize $L(\phi, \psi)$ while fixing the sum $\phi + \psi = 2\alpha$ we take $\phi = \psi = \alpha$.

Corollary. When D^{ij} is symmetric

$$r_{ab} = \min_{\phi(a)=1, \phi(b)=0} L(\phi, \phi). \quad \blacksquare$$

This minimum principle for resistances was known already to 19th century physicists, specifically Thomson (a.k.a. Kelvin), Maxwell, and Rayleigh: For more about this, see Doyle and Snell [5].

Having a straight minimum is a lot better than having a minimax, because now we can plug in any ϕ with $\phi(a) = 1, \phi(b) = 0$ and get an upper bound for r_{ab}, corresponding to a lower bound for T_{ab}. This method is a staple of electrical theory—the part of electrical theory that doesn't extend to non-time-reversible chains because it depends on the relation $L(f, g) = L(g, f)$.

For time-reversible chains there are also complementary methods for finding lower bounds for r_{ab}, and thus upper bounds for T_{ab}. These emerge from the minimum principle through the mystery of convex duality. In practice, though, it is generally conceptually simpler to work instead with the monotonicity law described in the next section. This monotonicity law extends to all chains, but sadly, for all we can tell thus far, its usefulness appears to get left behind.

9 Monotonicity

From the minimax characterization of commute times we immediately get the following:

Monotonicity Law. Commute times decrease monotonically when equilibrium interstate transition increase: Using barred and unbarred quantities to refer to two different Markov chains, if $D^{ij} \leq \bar{D}^{ij}$ for all $i \neq j$ then $\bar{T}_{ij} \leq T_{ij}$ for all i, j. ■

Actually it would be better to think of D and $\bar{D}$ here as referring to conformal classes of chains, rather than individual chains, because as we know D^{ij} and T_{ij} are conformal invariants.

This law holds for all chains, time-reversible or not. As we said above, for time-reversible chains this law can be used to get upper and lower bounds for commute times, and hence for hitting probabilities: This is discussed in great detail by Doyle and Snell [5].

Sadly, even though the law extends to the non-time-reversible case, its usefulness does not extend, at least not in any obvious way. How can this be? There seem to be a number of reasons.

First, for time-reversible chains, if we block transitions back and forth between states c, d, requiring the particle to remain where it is when it attempts to make such a transition, we get a new $\bar{D}$ dominated by the original D in the sense that $DD^{ij} \leq D^{ij}$ for $i \neq j$. Electrically speaking, blocking transitions between c and d amounts to cutting the wire between them. In the non-time-reversible case, this will change the equilibrium measure w^i and thereby destroy the relation $\bar{D}^{ij} \leq D^{ij}$ that we need for monotonicity.

Second, for time-reversible chains, it is simple and natural to introduce intermediate states. Electrically speaking, introducing a state between c and d amounts to dividing the 'wire' connecting c and d into two pieces, if only in our mind's eye. By combining this with the putting or taking of wires,

we can produce chains to bound T_{ab} above or below as closely as we please. And we can do this in such a way that our approximating chains are easy to analyze. Here lies the third apparent shortcoming of the non-time-reversible case: A seeming paucity of chains whose commute times are easy to compute.

So, of what use is this monotonicity law in the non-time-reversible case? That remains to be seem.

10 The obstruction to time-reversibility

Let M_{ij} be the expected time to reach j starting from i. Coppersmith, Tetali, and Winkler [**?**] showed that a Markov chain is time-reversible just if for all a, b, c

$$M_{ab} + M_{bc} + M_{ca} = M_{ac} + M_{cb} + M_{ba}.$$

And in this case the expected time to traverse a cycle of any length will be the same in either direction. Note that though the M_{ij}s themselves are not conformally invariant, these cycle sums are. For a cycle of length 2, the cycle sum is our best friend the commute time.

We always have

$$M_{ab} + M_{bc} + M_{ca} = \hat{M}_{ac} + \hat{M}_{cb} + \hat{M}_{ba}$$

(look at a long record of the chain backwards), so an equivalent condition is that for all a, b, c

$$M_{ab} + M_{bc} + M_{ca} = \hat{M}_{ab} + \hat{M}_{bc} + \hat{M}_{ca}.$$

This is true despite the fact that in general

$$\hat{M}_{ab} \neq M_{ba}.$$

So, why is this criterion true? A conformal class of chains is reversible just if our bilinear form $L(\phi, \psi)$ on $V = \{x^i | \sum_i x_i = 0\}$ is symmetric. Call two matrices *cohomologous* if they differ by a matrix of the form $B_{ij} = a_i - a_j$. To any bilinear form $\sum_{ij} u^i Z_{ij} v^j$ on V there corresponds a natural cohomology class of matrices

$$[\eta_{ij}] = [Z_{ij} - Z_{ji}],$$

which is to say, an antisymmetric matrix defined up to addition of a matrix of the form $B_{ij} = a_i - a_j$. Take for Z the Green's function, and call the resulting

$[\eta]$ the *resistive obstruction* of our Markov chain. The class $[\eta]$ vanishes just if η integrates to 0 around any cycle. Cycles of length 3 span the space of cycles, and integrating η around the cycle i, j, k gives the difference of traversal times above. So for time reversibility it is enough to require that traversal times agree forwards and backwards around any triangle. Indeed, this is more than enough, because triangles span the space of cycles in a very redundant way. To verify reversibility, it would suffice to check any basis for the space of cycles, e.g. only cycles of length 3 involving the fixed state r (the 'ground').

11 More to be said

The next step would be discuss how to use the knee-jerk mapping to make a chain time-reversible without changing its commute times. The knee-jerk method will produce the desired time-reversible chain whenever such a thing exists, but we still don't know if this is always the case. What we do know is that if it turns out that no suitable time-reversible chain exists, the knee-jerk method will produce a time-reversible chain whose commute times agree as well as possible with those of the original chain. (See Coppersmith et al. [2], Doyle [4] .)

Then we should discuss uniformization of Markov chains, whereby we prescribe a canonical representative chain within each conformal class (or in other words, we prescribe a canonical w to accompany a given N). This canonical chain extremizes the *seek time* or *Kemeny constant*, which is the expected time $M_{iw} = \sum_j M_{ij} w^j$ to reach a random state chosen according to the equilibrium distribution w: As Kemeny observed, M_{iw} doesn't depend on the starting point i. (Cf. Doyle [3].) The extremal chain is characterized by constancy of the expected time $M_{wj} = \sum_i w^i M_{ij}$ to hit i starting from equilibrium. (Note how this differs from the seek time, where the target point was random: Here it is the starting point that is random!) It's easy to write down the transition probabilities for this extremal chain. But, are they necessarily positive?

Beyond this lies the extension of this whole business to diffusion on surfaces, where we must renormalize hitting times because Brownian motion in dimension 2 never hits a given point. (Cf. Doyle and Steiner [6].) Now to uniformize we should extremize not Kemeny's constant, but a variant with a correction term involving the Gaussian curvature. Again, it is easy to write

down the extremizing metric, or rather the extremizing area measure, which is not a priori positive everywhere. For spheres, all the round metrics tie for the extremum. For tori, the flat metrics win. For higher genus surfaces, the winners are not hyperbolic surfaces, nor should they be, because having constant curvature is a local condition that doesn't know thick from thin. The canonical measure is sensitive to thickness in a conformally correct way. But is it a positive measure? If it isn't, could it still be good for something? Well, actually, it seems to give the so-called 'canonical metric' on a Riemann surface, so it looks like we're onto something.

References

[1] A. K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 574–586, Seattle, May 1989.

[2] D. Coppersmith, P. Doyle, P. Raghavan, and M. Snir. Random walks on weighted graphs, and applications to on-line algorithms. *Journal of the ACM*, 40:454–476, 1993.

[3] Peter G. Doyle. The Kemeny constant of a Markov chain, arXiv:0909.2636v1 [math.PR].

[4] Peter G. Doyle and Jim Reeds. The knee-jerk mapping, arXiv:math/0606068v1 [math.PR].

[5] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*. The Mathematical Association of America, 1984, arXiv:math/0001057v1 [math.PR].

[6] Peter G. Doyle and Jean Steiner. Hide and seek on surfaces and Markov chains.

[7] F. Kelly. *Reversibility and Stochastic Networks*. Wiley, 1979.

[8] J. F. C. Kingman. Markov population processes. *j. Appl. Prob.*, 6:1–18, 1969.