Complex Systems, Statistical Learning, and Pickles
Lecture 1 Notes for CSSS2010

Gregpry Leibon

June 15, 2010

Disclaimer: These are some notes to help with the lectures. They are **very rough** (especially the grammar); the exercises in particular come with **No Guaranty's**. (Of course, if you have the intestinal fortitude, then all the better!) Any corrections and feedback would be super-duper appreciated. I do not deal with basic aspects of the lecture already in [6], or very basic aspects of Markov chain theory (see [3]). The fundamental role of of commute times was introduced in [1]. The electric network theory underlying that work is beautifully described in [2]. One could redo this presentation from the point of view of the Green's function, this (and much more) can be found in [4] and [5].

A brief comment on the notation used here. Sates will be denoted with lower case letters like $i, j, k, l...$ or $a, b, c, d, e, ...$ A vector like v will denote a column vector, and when viewing v as a row vector we use v^{tr}. We will denote the matrix with a given vector v on the diagonal and zero otherwise as $[v]$. We let $\mathbf{1}$ denote the vector of ones, $\mathbf{0}$ denote the vector of zeros, δ_a will be the vector which is 1 at a and 0 otherwise, and we will let I denote the identity matrix.

1 Continuous time Markov chains

As we mentioned in [6], we can sometimes approximate the dynamics on a network by viewing the dynamics as a Markov chain. We have thought about the discrete case, where we defined that the transition form a state i to a state j occurs with probability P_{ij}. Even as a first order approximation, it is often useful to use continuous time chains, and our first order of business is to describe what such chains are. When a transition occurs, a continuous time Markov chain still transitions according to the probabilities of a discrete time chain that we will denote as P. We will assume that this underlying discrete chain is ergodic, which means that it is possible to get from any state to any other state. An ergodic discrete chain P has an equilibrium measure which we will denote as π^{tr}. Furthermore, when dealing with continuous chains is is very natural to assume $P_{ii} = 0$, and we will assume $P_{ii} = 0$ for our chains (see section 1, exercise 3).

What makes a continuous time Markov chain continuous time is that there is a positive expected time that one stays at state i before leaving, which we will denote as τ_i. The time spent at a state before leaving will be memoryless, hence the expected time before leaving will be exponentially distributed with rate $1/\tau_i$ (see section 1, exercise 1). The equilibrium measure for the continuous time chain must be weighted to account for the waiting time and is now $\omega_i = \tau_i \pi_i.$, where we restrict ourselves to τ_i such that this ω^{tr} is a probability measure (in other words $\omega^{tr}\mathbf{1} = 1$).

The key object to compute will be $P^t = P(X_t = j \mid X_0 = i)$. We will always use a superscript t when we refer to P^t and without the t we are referring to the transition matrix of the underlying discrete chain P. The instantaneous rate of transition satisfies

$$\frac{d}{dt}P_t\bigg|_{t=0} = -\Delta_\omega$$

where

$$\Delta_\omega = [\tau]^{-1}(I - P),$$

and we will call this the chain's *Laplacian*. The system being Markov mean this is the chains generator and $P^t = e^{-t\Delta_\omega}$ (see section 1, exercise 2).

Exercises:

1. Here we explore the notion of memoryless.

   (a) Given $X_{t_0} = i$, let T be the first time one transitions from i. Explain why memoryless should mean for $s > t_0$ we have $P(T > (t+s) \mid T > s) = P(T > t)$.

   (b) Use this to show $\frac{dP(T>t)}{dt} = CP(T > t)$.

   (c) Prove that $P(T > t) = e^{-\frac{t}{\tau_i}}$.

2. We let X_t denote the state we are at time t. Four finite chains Markov can be defined as

$$P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i). \qquad \text{(Memoryless)}$$
$$P(X_{t_1} = j \mid X_{t_0} = i) = \textstyle\sum_k P(X_{t_2} = j \mid X_{t_1} = k)P(X_{t_1} = k \mid X_{t_0} = i). \qquad \text{(ChapmanKolmogorov)}$$

   (a) Prove the power series $e^z = \sum_{n=1}^{\infty} \frac{z^n}{n!}$ converges when a matrix A is plugged in. This allows us to define e^A.

   (b) Prove $P^t = Ce^{tA}$ solves $\frac{d}{dt}P^t = AP^t$ with $P^0 = C$.

   (c) Prove that the definition of memoryless implies $\frac{d}{dt}P^t = -\Delta_\omega P^t$ with $P^0 = I$ (hence $P^t = e^{-t\Delta_\omega}$).

3. Here we dwell on the $P_{ii} = 0$ restriction.

   (a) Does it make sense for a continuous time Markov chain X_t to have $P_{ii} > 0$?

   (b) If so, is such a chain equivalent to a unique chain with $P_{ii} = 0$?

1.1 Markov chain examples

Toy Chain: It is nice to have a Toy example to play with. Our toy example will be derived from the weights on the edges of the undirected graph in figure 1. We will call the matrix where W_{ij} is the weight on the edge connection states i and j the *weight matrix* or the *conductance matrix*. Notice W is symmetric. We produce the underlying discrete chain via $P = [W\mathbf{1}]^{-1}W$. Our chain was intentionally endowed withe some geometry. The low weight edges (where $0 < W < 20$) have been colored yellow, while the high weight edges (where $80 < W < 100$) form three connected components, denoted by the set of red edges, the set of green edges, and the set of blue edges. The states in these components are also colored, with the blue component in light and dark blue, with dark blue it's little tail. The Black dots are hard to get to from the main components, since at a vertex with both high and low weight edges the higher weight edges will correspond to the directions with a comparatively high probability of transition. In what follows, we will give this chain various waiting time choices.

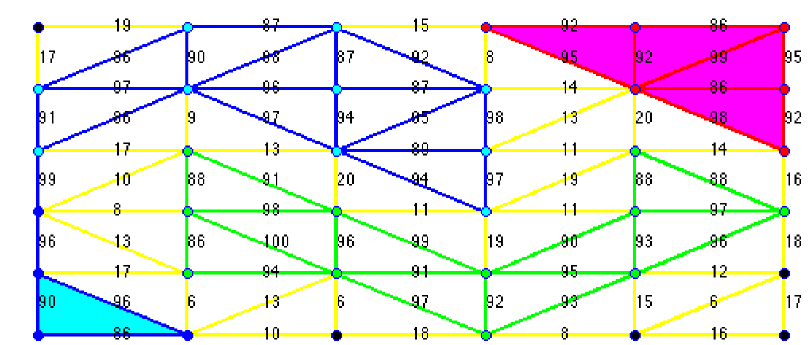

Figure 1: Our toy chain.

Space of Mathematics: We model the process of exploring the space of mathematics by collecting all of the roughly 18,000 mathematics pages on Wikipedia's *list_of_mathematics_articles* [1] (excluding *list_of_*, and other clearly inappropriate pages, see [7]). We created a not necessarily symmetric weight matrix U which is zero unless there is a link on page i to page j, in which case $U_{ij} = 1$. We also included the *list_of_mathematics_articles* page, and linked every one to it with a lower weight (I chose $1/5$). This helped capture the notion of "using a search engine", and also makes the chain determined by $P = [U\mathbf{1}]^{-1}U$ clearly

[1]Special thanks to Kyle Lad and Miles Kenyon for collecting this data and examining its structure.

ergodic. This chain is a first order approximation of a web like exploration (see section 1.1, exercise 1). We assume $\tau = 1$.

Exercises:

1. In the space of mathematics example, we approximated web exploration as a Markov chain.
   (a) Explain the effect of the back button and the use of a search engine in web exploration.
   (b) How might you improve our model?
   (c) We assume $\tau = 1$. How might you make this assumption more realistic?

1.2 Conformally invariant properties

We define different choices of τ with the same P as *conformal changes* and any properties dependent only on P will be called *conformally invariant*. For example, if we view each transition as a step we can think about the number of steps we take as opposed to time, and look at the number of steps taken at time t as $N(t)$. Concepts based on number of steps to do something are of course conformally invariant, but sometimes concepts involving time turn out to be conformally invariant. For example, from discrete Markov chain theory we know that starting at a that the expected number of steps until reaching a again is $1/\pi_a$ (see section 1.2, exercise 1). If we define the first return time to be the **time** it takes starting at a to reach a again, then:

Conformal Invariance of First Return Time: The expected first return time at a is conformally invariant, and hence equal to $1/\pi_a$.

proof: Denote the return time random variable as as R_a. Run the chain staring at i. Look at the time t_1 of the first return. You do this again and find its time t_2, and so on. By the law of large numbers

$$E(R_a) = \lim_{T\to\infty} \frac{\sum t_i}{M_a(N(T))},$$

where $M_a(N)$ is the number of returns until the Nth step and $N(T)$ is the number of steps up to time T. Now since $\sum \omega_i = 1$ let us observation $N(T)$ asymptotically depends only on the underlying discrete structure. Hence, since our t_i cover the whole interval once (except a negligibly little piece at the end),

$$E(R_a) = \lim_{T\to\infty} \frac{T}{M_a(N(T))},$$

and hence involves only the conformal structure as claimed. **Q.E.D**

There is a conformally invariant version of the generator Laplacian called the *Kirchhoff Matrix*, defined as

$$K = [\omega]\Delta_\omega.$$

As in [2], the Kirchhoff Matrix allows us to bring in various electrical analogies as the relationship between charge distribution ρ and electric potential V is given by

$$KV = \rho.$$

There is also a conservation of charge that requires

$$\sum \rho = 0.$$

For chains determined by a symmetric weight matrix W, this electrical analogy is quite literal (see [2]), but we will use the same language for arbitrary chains.

Exercises:

1. Follow these steps to prove for a discrete chain that $E(R_a) = 1/\pi_a$. To do so we define the mean first passage time matrix to be the matrix such that M_{ab} is the expected time to get from a to b.

   (a) Denote r the vector with i^{th} entry $E(R_a)$. By taking a step show

   $$M = PM + \mathbf{1}\mathbf{1}^{tr} - [r].$$

   (b) From the previous problem $(I-P)M = \mathbf{1}^{tr}\mathbf{1} - [r]$., now apply π to both sides to show $E(R_a) = 1/\pi_a$.

   (c) Show M is **not** conformally invariant.

2. For a discrete time chain with $P_{ii} \geq 0$, one might chose to define conformally equivalent chains as those sharing the same Kirkoff Matrix K.

   (a) Describe why this is a good idea.

   (b) Show every such discrete with $P_{ii} \geq 0$ chain is conformally equivalent to a unique chain with $P_{ii} = 0$.

3. The Green's Function. Let

   $$G = (K + \pi\pi^{tr})^{-1} - \mathbf{1}\mathbf{1}^{tr}$$

   (a) Show $KGK = K$, $G\pi = 0$, and $\pi^{tr}G = \mathbf{0}^{tr}$, and that these conditions uniquely determine G.

   (b) Show $V = G(\rho)$ solves $KV = \rho$ for ρ such that $\sum \rho = 0$.

   (c) Show G can be expressed as

   $$G = \left(\sum_{n=1}^{\infty} (P - P^{\infty})^n \right) [\pi]^{-1}$$

   where $P^{\infty} = \mathbf{1}\pi^{tr}$, and explain the notation.

2 Expected commute time

Let us define the *Commute Time* as T_{ij} as the random variable telling the time that it take to start at i hit j and return to i.

Conformal Invariance of the Expected Commute Time: The expected commute time depends only on the underlying discrete structure.

proof: This is a schollium of the Conformal Invariance of the First Return Time Lemma. In this case, t_1 is the first commute back to i via j. Now you begin you second commute and find its time t_2, and so on. Just as was the case for the return times, our t_i cover the whole interval once (except a negligibly little piece at the end), implying that $E(T_{ab})$ involves only the underlying conformal structure. **Q.E.D**

Inspired by the idea of plugging a battery onto our network to a positive unit charge a a negative unit charge at b, we define V_a^b as

$$KV_a^b = \delta_a - \delta_b.$$

In the wold of electric networks, $V_a^b(b) - V_a^b(a)$ is the known as the *effective resistance*, but in complete generality we have:

Commute Theorem:

$$E(T_{ab}) = V_a^b(b) - V_a^b(a)$$

Comment: There is an alternate, perhaps simpler proof using the Green's Function (see section 2, exercise 4). But this proof give us the opportunity to introduce some useful ideas (from potential theory and renewal theory) that we will make use of later.

proof: By the Conformal Invariance Lemma we only need to prove this in the discrete case. Now in the discrete world

$$q(x) = \frac{V_a^b(x) - V_a^b(a)}{V_a^b(b) - V_a^b(a)}$$

has lovely interpretation as $p(x)$: the probability of starting at of reaching b before hitting a. To see this it is useful to note $q(x)$ is *harmon function*. A harmonic function is a function f which satisfies $\Delta f = 0$ off a specified set call the *boundary*. Hence our q is clearly harmonic with boundary $\partial = \{a,b\}$. Notice that away from ∂, that if I take one step then $p(x)$ is equal to the average of its neighbors. Hence both p and q are harmonic with $\partial = \{a,b\}$ and furthermore $p(a) = q(a) = 0$ and $p(b) = q(b) = 1$. Our claim is that this forces $p = q$ as hoped for. To accomplish this, we look at the function $f = p - q$. By the linearity of Δ this function is harmonic and zero on the boundary. So $p - q = 0$ everywhere follows from the following principle:

Maximal Principle: A harmonic function is constant or takes on its maximum on the boundary.

Proof Maximum Principal: If the maximum is in the interior, then the function at this point being the average of its neighbors forces all it's neighbors to be the same value. So it is constant everywhere (recall our chain is ergodic, hence all point are connected by a sequence of transitions). **Q.E.D Maximum Principal**

Now we need another concept *the escape probability*. Notice $(\Delta_\pi p)(a) = p_e$ is the probability that we escape from a, in other words leaving a we reach b before returning to a. Since $KV_a^b = \delta_a - \delta_b$ we have $\Delta_\pi q(a) = \frac{1}{\pi_a(V_a^b(b) - V_a^b(a))}$. So from the equality of p and q, we now see that

$$V_a^b(b) - V_a^b(a) = \frac{1}{\pi_a p_e}.$$

We are done if we can demonstrate:

Lemma: $E(T_{ab}) = \frac{1}{\pi_a p_e}$

Proof Lemma: Notice that starting at a I can view my voyage back to a via b as a sequence of M returns to a, each involving $R_{a,i}$ steps, together with one last one where I go via b. Hence $T_{a,b} = \sum_{i=1}^M R_{a,i}$ and

$$\begin{aligned}
E(T_{a,b}) &= E(\textstyle\sum_{i=1}^M R_{a,i}) && \text{(just observed)}\\
&= E(E(\textstyle\sum_{i=1}^M R_i \mid M = m)) && \text{(basic expectation property)}\\
&= \textstyle\sum_{i=1}^m E(\sum_{i=1}^m R_i) P(M = m) && \text{(def. expected val.)}\\
&= \textstyle\sum_{i=1}^m E(\sum_{i=1}^m R_i) p_e (1-p_e)^{m-1} && \text{(escaped m-1 times)}\\
&= \textstyle\sum_{i=1}^m m E(R_i) p_e (1-p_e)^{m-1} && \text{(linearity exp. val.)}\\
&= p_e E(R_i) \textstyle\sum_{i=0}^m m(1-p_e)^{m-1} && \text{(rearrangement)}\\
&= \tfrac{p_e}{\pi_a} \textstyle\sum_{i=0}^m m(1-p_e)^{m-1} && \text{(expected return time, Sec.1.2,Ex.1)}\\
&= \tfrac{p_e}{\pi_a} \tfrac{1}{p_e^2} && \text{(geometric series, Sec.2,Ex.1)}\\
&= \tfrac{1}{\pi_a p_e}
\end{aligned}$$

Q.E.D Sub-lemma and Q.E.D

Exercise:

1. In the above proof we used the single most important fact about infinity: the convergence of the geometric series.

   (a) Prove the $\sum_{i=0}^\infty r^n = \frac{1}{1-r}$.

   (b) Take the derivative to prove $\sum_{i=0}^\infty n r^{n-1} = \frac{1}{(1-r)^2}$.

2. We used one of the very basic fact from probability in our proof of the Commute Theorem , namely that $E(E(X \mid Y)) = E(X)$.

   (a) Generalize our use of this fact to prove the Wald's Theorem

$$E\left(\sum_{i=1}^M X_i\right) = E(M)E(X)$$

(b) Use $E(E(X \mid Y)) = E(X)$ to show if X and Y are independent then $E(XY) = E(X)E(Y)$.

(c) Use $E(E(X \mid Y)) = E(X)$ to show if X and Y are independent then $Var(X+Y) = Var(X) + Var(Y)$.

3. (hard, see [**?**] for a lovely proof) We introduced the notion of escape probability p_e in order to prove the Commute Theorem. For an infinite graph we can escape to ∞ from a provided we leave a and never return.

   (a) Prove that when random walking on the square lattice in R^2 that $p_e = 0$.

   (b) Prove that when random walking on the square lattice in R^3 that $p_e > 0$.

4. Alternate proof of the Commute Theorem using the Green's Function.

   (a) Notice that $T_{ab} = M_{ab} + M_{ba}$ where M is the first passage time section 1.2 exercise 1.

   (b) Prove that $M_{ab} = G_{ab} - G_{bb}$

   (c) Use section 1.2 exercise 3 part (b) to prove $G_{ab} - G_{aa} = V_b^a(a)$, and prove $T_{ab} = V_b^b(b) - V_b^a(a)$.

3 Green's embedding

In this section we demonstrate that there is a natural embedding of our chains sates into Euclidean space such that the distance between the states is in fact $\sqrt{E(T_{ab})}$. We call this distance the *commute distance*. We will embed our states by route of a geometry on the function space on our nodes. It is not the simplest geometry on this function space which is given by the what is usually called the L^2 structure on the functions, namely

$$< f, g >_\omega = f^{tr}[\omega]g.$$

To find the correct geometry we need Green's Identity:

Green's Identity:

$$< f, \Delta_\omega f >_\omega = \frac{1}{2}\sum_{i,j} \pi_i P_{ij}(f_i - f_j)^2$$

proof:

$$\begin{aligned}
< f, \Delta_\omega f >_\omega &= \textstyle\sum_{i,j} f_i \omega_i \frac{1}{\tau_i}(I_{ij} - P_{ij}) f_j && \text{(def.)} \\
&= \textstyle\sum_{i,j} f_i \pi_i (I_{ij} - P_{ij}) f_j && (\omega_i = \tau_i \pi_i) \\
&= \tfrac{1}{2}\left(\textstyle\sum_i \pi_i f_i^2 - 2\sum_{i,j} f_i \pi_i P_{ij} f_j + \sum_j \pi_j f_j^2\right) && \text{(foil)} \\
&= \tfrac{1}{2}\left(\textstyle\sum_{i,j} \pi_i P_{ij} f_i^2 - 2\sum_{i,j} f_i \pi_i P_{ij} f_j + \sum_j \pi_j f_j^2\right) && \text{(rowsum P is 1)} \\
&= \tfrac{1}{2}\left(\textstyle\sum_{i,j} \pi_i P_{ij} f_i^2 - 2\sum_{i,j} f_i \pi_i P_{ij} f_j + \sum_{i,j} \pi_j P_{i,j} f_j^2\right) && (\pi \text{ is } P \text{ equi. vec.}) \\
&= \tfrac{1}{2}\textstyle\sum_{i,j} \pi_i P_{ij}(f_i^2 - 2 f_i f_j + f_j^2) && \text{(factor)} \\
&= \tfrac{1}{2}\textstyle\sum_{i,j} \pi_i P_{ij}(f_i - f_j)^2 && \text{(unfoil)}
\end{aligned}$$

Q.E.D

From the Green's Identity we see that

$$< f,g >_{Dir} = < f,\Delta_\omega g >_\omega .$$

is an non-negative, bilinear form called the *Dirichlt form*, and it is this form that determines the correct geometry. Notice from Green's Identity we have that the Dirichlet form is a well defined positive form on $W = R^N / < const >$ where $< const >$ is the span of the constant functions. Furthermore, we can also symmetrize the Diriclhet form without effecting the norms of functions, and we define

$$< f,g >_{SymDir} = \frac{1}{2}(< f,g >_{Dir} + < g,f >_{Dir}),$$

which is now seen to be a Euclidean inner product on on W. We now prove that need embedding is into $(W, < \cdot,\cdot >_{SymDir})$ and given by

$$Green(a) = V_g^a$$

for a fixed choice of ground state g.

Green's Embedding:

$$||Green(a) - Green(b)||^2_{SymDir} = E(T_{ab})$$

proof:

$$
\begin{aligned}
&||Green(a) - Green(b)||^2_{SymDir} = \left|\left|V_g^a - V_g^b\right|\right|^2_{SymDir} && \text{(def.)}\\
&= \left|\left|V_g^a - V_g^b\right|\right|^2_{Dir} && \text{(unsym.)}\\
&= \left|\left|V_a^b\right|\right|^2_{Dir} && \text{(} \Delta \text{ linearity, see 3ex.3)}\\
&= < V_a^b , \Delta V_a^b >_\omega && \text{(def. } Dir\text{)}\\
&= (V_a^b)^{tr} K V_a^b && \text{(def. } K \text{)}\\
&= (V_a^b)^{tr}(\delta_a - \delta_b) && \text{(def. } V_a^b\text{)}\\
&= V_a^b(a) - V_a^b(b) && \text{(evaluate)}\\
&= E(T_{ab}) && \text{(Commute Theorem)}
\end{aligned}
$$

Q.E.D

Exercises:

1. Using the Green's function from section 1.2 exercise 3, to form an explicit embedding, by showing that $||Fe_a - Fe_b||^2_{StandEucl} = E(T_{ab})$ where $F = \sqrt{(K + K^{tr})/2}G$.

2. Look up Green's identity from calculus. Explain why the above the Green's Identity is called the Green's Identity.

3. Superposition

   (a) Show $V_a^b = V_a^g - V_b^g$.

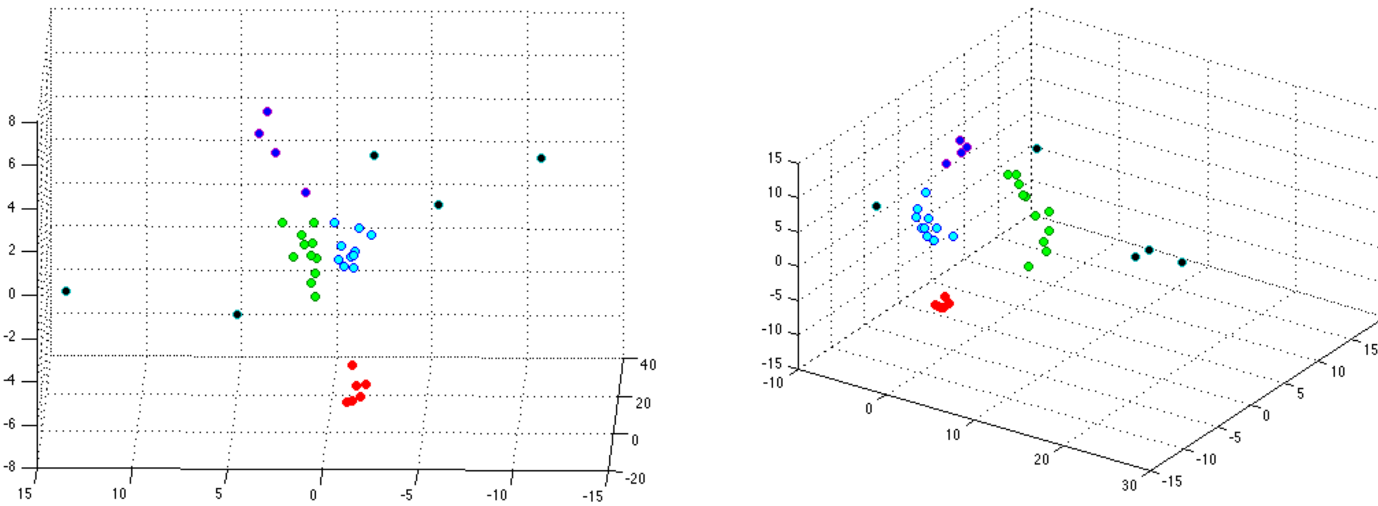

Figure 2: On the left we have the the Green's embedding of the toy chain. On the right we use weighted MDS.

(b) Given $\sum \rho = 0$ and a fixed ground state g, prove there unique constants c_a such that $\rho = \sum_{a \neq g} c_a(\delta_a - \delta_g)$.

(c) Prove the solution to $KV = \rho$ is $V = \sum c_a V_a^g$.

3.1 Green's embedding examples

In this section, we Green's embed our examples.

Toy Example: On the left in figure 2 we use the commute distance $d_{Com}(a,b) = \sqrt{E(T_{ab})}$ and MDS to witness the Green's embedding. Notice the black dots and the hard to reach dark blue dots are indeed very far away, hence have a large commute times to other points as intuitively expected. The core geometry of the three main components is clearly exposed in the center. If one wants to emphasize the core geometry one could weight the MDS as in the section 3.2.

Space of Mathematics: We find that the commute time between the *Fundamental Theorem of Calculus* and *Stokes Theorem* is roughly 5287, while the commute time between the *Fundamental Theorem of Calculus* and the *Gauss Bonnet Theorem* is roughly 34809. In figure 3, we see the embedding of some of our favorite theorems using the distance $d_{Com}(a,b) = \sqrt{E(T_{ab})}$.

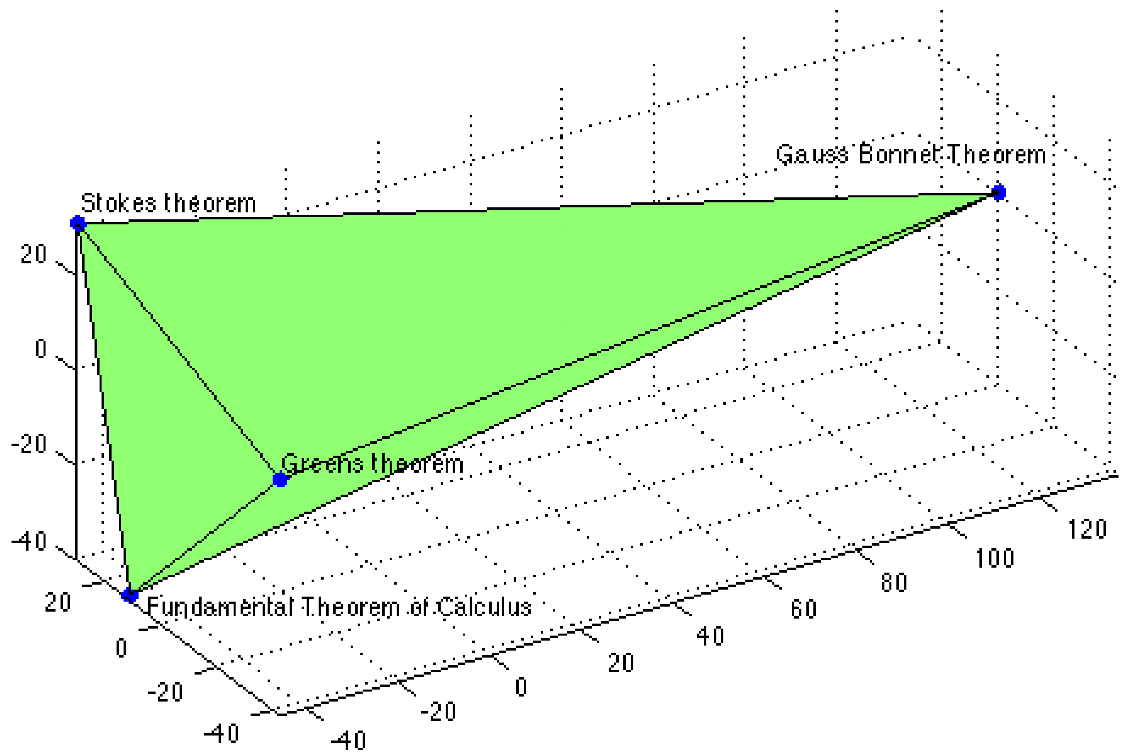

Figure 3: Here we look at few points in the Space of Mathematics using the Green's embedding.

3.2 Incorporating the non-conformal

By the Conformal Invariance of the Expected Commute Time, we see that the the Green's embedding is conformally invariant. But the conformal part of the geometry, stored in ω, is not wise to ignore. In particular, we can use ω to help us emphasize the part of the chain were the action is. One way to accomplish this is to use the *weighted MDS*, which replaces the minimization of a usual objective function like

$$stess = \sum_{i>j} (d_{Euc}(f(x_i), f(x_j)) - d(x_i, x_j))^2$$

with a version weighted with positive constants W_{ij}, like

$$stessWeighted = \sum_{i>j} W_{i,j} (d_{Euc}(f(x_i), f(x_j)) - d(x_i, x_j))^2 .$$

To emphasize the part of he chain where we spend more time we can use

$$W_{ij} = g(\omega_i)g(\omega_j)$$

with g an increasing function.

Toy Example with weighted MDS: On the right in figure 2, we use $d_{Com}(a,b) = \sqrt{E(T_{ab})}$ and the weighted MDS with weights $W_{ij} = \sqrt{\omega_i}\sqrt{\omega_j}$ to view our chain. Notice the embedding does appear to better capture the components where we spend more time.

Exercises:

1. Our Green's embedding is into Euclidean space, when the distances are Euclidean MDS (using the above *stress*) agrees with PCA.
   (a) Using the explicit embedding in section 3 exercise 1 to confirm this on an example.
   (b) Prove if the distances are Euclidean that MDS (using the above *stress*) agrees with PCA.
2. Weighted PCA.
   (a) Look up or simply develop a weighted version of the PCA.
   (b) Use the equillibrium measure and a weighted PCA to embed your favorite chain from its Green's embedding.
   (c) Compare the results to the MDS embedding and the weighted MDS embeddings.

4 References

The use of commute time to understand Markov chains was introduced in [1]. *The Electrical Resistance Of A Graph Captures Its Commute And Cover Times*, by Ashok K. Chandra , Prabhakar Raghavan , Walter L. Ruzzo , Roman Smolensky, and Prasoon Tiwari; which is available at http://www.cs.washington.edu/homes/ruzzo/papers/resi

The single best reference is *Random Walks and Electric Networks*, by Peter G. Doyle and J. Laurie Snell; which is available at http://arxiv.org/abs/math/0001057

References

[1] 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*, http://www.cs.washington.edu/homes/ruzzo/papers/resist.pdfDBLP

[2] Peter G. Doyle and J. Laurie Snell, *Random Walks and Electric Networks*, http://arxiv.org/abs/math/0001057

[3] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability*, http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/Chapter11.pdf

[4] Peter G. Doyle, Gregory Leibon, and Jean Steiner *Conformal geometry of Markov chains*, pickle.pdf as posted on CSSS10 Wiki

[5] Peter G. Doyle, Gregory Leibon, and Jean Steiner *Markov chains in a shoebox* , shoe.pdf as posted on CSSS10 Wiki

[6] Gregory Leibon, Scott Pauls, Dan Rockmore, and Robert Savell *Statistical Learning for Complex Systems an Example-driven Introduction* , book.pdf as posted on CSSS10 Wiki

[7] Kyle Lad *Python code to mine the Space of Mathematics* , SpaceOfMath.zip as posted on CSSS10 Wiki