BOUNDS ON BIASED AND UNBIASED RANDOM WALKS

DANIEL PARRY

Abstract. We analyze several random random walks on one-dimensional lattices using spectral analysis and probabilistic methods. Through our analysis, we develop insight into the pre-asymptotic convergence of Markov chains.

1. Introduction

The pre-asymptotic convergence of Markov chains is a relatively new field of study — only two or three decades old — and is still an active area of research. One example of a pre-asymptotic behavior is the "cutoff phenomenon" explored by Diaconis and his collaborators. A Markov chain has a cutoff if it remains far from stationary for a long period, after which it converges within a small number of iterations. As his most famous example, Diaconis showed that seven shuffles is enough to randomize the order of a deck of cards, but after six shuffles the card order is still far from uniformly randomized. Though many examples have been analyzed, in general the cutoff phenomenon is still not well understood [1].

Our goal in this paper is to explore the cutoff phenomena for some random walks on one-dimensional lattices. After reviewing some facts about discrete Markov chains in general, we describe spectral and probabilistic bounds that describe their convergence.

We pick two examples to study the convergence to stationary, the biased random walk and the unbiased random walk on a bounded one-dimensional lattice. Intuitively, we expect the biased random walk to have a cutoff, because the distribution for a similar walk on an unbounded domain is a nearly-Gaussian "blob" forever heading towards one of the "'boundaries"'; but the stationary distribution for the bounded walk is exponential. The Gaussian blob will remain as long as a substantial amount of probability does not interact with the boundary conditions. If N is the size of the domain, it should take approximately $N/|\mu| - O(\sqrt{N})$ steps before the Gaussian blob starts to substantially interact with the boundaries, where μ represents the expected motion at each step. At the point when the blob reaches the boundary, it should gradually change over to its final distribution. For large N, the total number of steps for the chain to converge will be $N/|\mu| + o(N)$. But the total variation distance to the stationary distribution will also be arbitrarily near one for $o(N)$ steps, and so there is a cutoff.

The unbiased walk behaves altogether differently. If we start an unbiased random walk from the middle of a domain of N points, it will take $O(N^2)$ steps before much probability reaches the boundary, and thus we expect to take $O(N^2)$ steps to

This research was supported by the Summer Undergraduate Research Experience (SURE) program at the Courant Institute.

converge. But because the probability smoothly diffuses toward a uniform distribution rather than advecting toward a final distribution with probability concentrated in a small area, the convergence is much more smooth.

2. Preliminaries

Definition 2.1. A discrete-time Markov chain is a sequence of random variables X_k in which each state depends only on the previous state, i.e., $P(X_{k+1}|X_k, X_{k-1}, \ldots) = P(X_{k+1}|X_k)$. For a time invariant Markov chain on states 0 through N, there is an associated matrix recurrence

$$\pi_{k+1} = \pi_k P,$$

where $P(X_k = j)$ is given by the jth component and $p_{i,j} = P(X_{k+1} = j|X_k = i)$. The matrix P is called the *transition matrix*. Note that P is row stochastic ($\sum_j p_{i,j} = 1$ and $p_{i,j} \geq 0$, $0 \leq i, j \leq N$) [7].

We will only consider Markov chains that are *reversible*, *irreducible*, and *aperiodic*. These properties are as follows [6]:

Definition 2.2. A Markov chain with transition matrix P is *reversible* if there exists some distribution π^* (called the *stationary distribution*) such that:

$$\pi_j^* p_{i,j} = \pi_j^* p_{j,i}.$$

Definition 2.3. A Markov chain is *irreducible* if for any states i, j there is an n such that $(P^n)_{i,j} > 0$.

Definition 2.4. A Markov chain is *aperiodic* if there is an n such that for all $M \geq n$, $0 < i \leq N$, $(P^M)_{i,i} > 0$.

Definition 2.5. A Markov chain that is irreducible and aperiodic is *ergodic*.

We now describe the convergence of finite Markov chains in terms of the eigendecomposition of the transition matrix. In particular, we will show that every ergodic reversible Markov chain has a Jordan form with eigenmatrix V and largest nonunitary eigenvector λ_2, such that:

$$\begin{aligned} V^T \Pi V &= I \\ \|\pi_k - \pi^*\|_1 &\leq \kappa_1(V)|\lambda_2^k|. \end{aligned}$$

Theorem 2.6. *For any reversible Markov chain P, there exists an eigenvector matrix V such that*

$$V^T \Pi V = I,$$

where Π is the matrix with the stationary distribution down its diagonal.

Proof. By reversibility, $\pi_j^* p_{i,j} = \pi_j^* p_{j,i}$ for $0 \leq i, j \leq N$. In matrix notation, $\Pi P = P^T \Pi$ or $\Pi P \Pi^{-1} = P^T$. Define $Q = \Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}}$; because Q and P are similar, the eigenvalues of P are the same as those of Q [3]. Also, Q is symmetric:

$$\begin{aligned} Q &= \Pi^{-\frac{1}{2}}(\Pi P \Pi^{-1})\Pi^{\frac{1}{2}} \\ &= \Pi^{-\frac{1}{2}} P^T \Pi^{\frac{1}{2}} \\ &= (\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}})^T \\ &= Q^T. \end{aligned}$$

By symmetry ($Q = Q^T$), Q has an orthonormal Jordan form ($Q = W\Lambda W^T$, where $W^T W = I$). Define the Jordan form $P = V\Lambda V^{-1}$ by

$$Q = W\Lambda W^{-1} = \Pi^{\frac{1}{2}} V\Lambda V^{-1}\Pi^{-\frac{1}{2}} = (\Pi^{\frac{1}{2}}V)\Lambda(\Pi^{\frac{1}{2}}V)^{-1}.$$

By inspection, $W = \Pi^{\frac{1}{2}}V$ and $V = \Pi^{-\frac{1}{2}}W$. This gives

$$V^T\Pi V = (\Pi^{-\frac{1}{2}}W)^T\Pi(\Pi^{-\frac{1}{2}}W) = W^T W = I.$$

□

The next theorems will link ergodicity to convergence. Ultimately, these theorems will lead to a second result.

Theorem 2.7. *All stochastic matrices have an eigenvalue of one and all other eigenvalues are of most one in magnitude.*

Proof. Suppose P is an arbitrary stochastic matrix. Any eigenvalue λ of P must lie in a union of Gershgorin discs [3]:

$$\lambda_i \in \bigcup_i \left\{ z \in \mathbb{C} : |z - p_{ii}| \leq \sum_{i \neq j} |p_{ij}| \right\}.$$

Because P is stochastic ($\sum_i p_{i,j} = 1$, $p_{i,j} \geq 0$), $\sum_{i \neq j} |p_{ij}| = 1 - p_{ii}$. By the triangle inequality, all eigenvalues lie in the disc $|z| \leq 1$. Furthermore,

$$Pe = e \quad \text{where} \quad e = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix}.$$

□

Theorem 2.8. *Suppose P is a transition matrix for some ergodic finite Markov chain. Then one is an eigenvalue of multiplicity one for all P. All other eigenvalues of P are strictly less than 1.*

Proof. If P is ergodic, then there is an n such that for all $k \geq n$, $(P^k)_{i,j} > 0$ for all $0 < i, j \leq N$, i.e. P^k is positive. Perron's theorem [3] states that a positive matrix has a unique dominant eigenvalue equal to the spectral radius. Since P is stochastic, this eigenvalue is one. □

Theorem 2.9. *Let P be a transition matrix of an ergodic Markov chain with initial distribution π_0. Then*

$$\lim_{k \to \infty} \pi_k = \pi^*.$$

Proof. Because P is ergodic it has a unique dominant eigenvalue at 1 with a corresponding row eigenvector w (assume w is normalized such that $\|w\|_1 = 1$) and column eigenvector e. Therefore P has a spectral decomposition $P = ew + J$, where

$J = (I - ew)P(I - ew)$ has spectral norm strictly less than 1. Therefore,

$$\begin{aligned}
\lim_{k\to\infty} \pi_k &= \lim_{k\to\infty} \pi_0 P^k \\
&= \lim_{k\to\infty} \pi_0 (ew + J)^k \\
&= \lim_{k\to\infty} \pi_0 (ew)^k + \pi_0 J^k \\
&= \lim_{k\to\infty} \pi_0 ew + \pi_0 J^k \\
&= \lim_{k\to\infty} w + \pi_0 J^k \\
&= w.
\end{aligned}$$

Therefore the chain converges to the unique stationary distribution $w = \pi^*$. □

Theorem 2.10 ([7]). *Given an ergodic Markov chain with transition matrix P and Jordan form $P = V\Lambda V^{-1}$, let λ_2 denote the largest eigenvalue other than one and $\kappa_1(V)$, the 1 norm condition number for V. Then*

$$\|\pi_k - \pi^*\|_1 \leq \kappa_1(V)\lambda_2^k,$$

Proof. The theorem is an extension of the argument from Theorem 2.9. Define $A = P - ew = V\hat{\Lambda}V^{-1}$. Then

$$\|\pi_k - \pi^*\|_1 = \|\pi_0 A^k\|_1 = \|V\hat{\Lambda}^k V^{-1}\|_1.$$

where

$$\hat{\Lambda}^k = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_{22}^k \end{bmatrix}.$$

Write the eigenvalues of Λ_{22} as $1 > |\lambda_2| \geq |\lambda_3| \geq \ldots$. Then we have

$$\|\pi_k - \pi^*\|_1 = \|V\Lambda^k V^{-1}\|_1 \leq \|V\| \, \|\Lambda\|^k \, \|V^{-1}\| = \kappa_1(V)\lambda_2^k.$$

□

3. Random Walk on a Lattice

Our main example is a random walk on $N + 1$ points. For this example we compute the eigendecomposition of the transition matrix P to estimate the bound $\kappa_1(V)|\lambda_2^k|$. The pre-asymptotic convergence of this random walk will differ depending on whether the walk is *biased* or *unbiased*, and we will treat these two cases in Section 4 and Section 5, respectively.

Definition 3.1. The *bounded random walk* has a transition matrix P of size $(N + 1) \times (N + 1)$ given by

$$P = \begin{bmatrix} p_s + p_l & p_r & 0 & \cdots & 0 \\ p_l & p_s & p_r & \cdots & 0 \\ 0 & p_l & p_s & p_r & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & p_l & p_s + p_r \end{bmatrix}$$

where $p_r, p_l, p_s > 0$, $p_l \geq p_r$ and $\sum p_i = 1$.

We can transform a bounded random walk into a walk such that $p_r \geq p_l$ by flipping the indices. This is equivalent to saying $\tilde{P} = QPQ^T$ where Q is the permutation matrix with ones on the anti-diagonal. Therefore, every theorem can be re-proven for $p_r \geq p_l$.

Theorem 3.2. *The bounded random walk is ergodic.*

Proof. Note that $(P^k)_{ii} \geq p_s(P^{k-1})_{ii}$ and $P^0 = I$. By induction, $(P^k)_{ii} \geq p_s^k > 0$, so the chain is aperiodic.

Similarly, we can show by induction on $|i-j|$ that

$$P_{ij}^{|j-i|} = \begin{cases} p_r^{j-i} > 0, & j \geq i \\ p_l^{i-j} > 0, & j \leq i. \end{cases}$$

Therefore the chain is also irreducible. □

Because the bounded random walk is ergodic, we know by Theorem 2.10 that it converges geometrically to a unique stationary distribution. In the remainder of this section, we prove that the stationary distribution is $\pi_k^* \propto (\frac{p_r}{p_l})^k$ and

$$\|\pi_k - \pi^*\| \leq N \left(\frac{p_r}{p_l}\right)^{-N/2} \left(p_s + 2\sqrt{p_r p_l}\cos\frac{\pi}{N+1}\right)^k. \tag{1}$$

Theorem 3.3. *The eigenvalues for the bounded random walk are $z = 1$ and $z = 2\sqrt{p_r p_l}\cos(\frac{\pi m}{N+1}) + p_s$ for $0 < m < N+1$.*

Proof. Translate $\pi_r P = z\pi_r$ into the system of difference equations:

$$\pi_{r,0}(p_s + p_l - z) + \pi_{r,1}p_l = 0, \qquad k = 0 \tag{2a}$$

$$p_r\pi_{r,k-1} + \pi_{r,k}(p_s - z) + \pi_{r,k+1}p_l = 0, \qquad 0 < k < N \tag{2b}$$

$$\pi_{r,N}(p_s + p_r - z) + \pi_{r,N-1}p_r = 0, \qquad k = N. \tag{2c}$$

Equation (2b) is a homogeneous, constant-coefficient difference equation. Looking at (2b) alone and ignoring the boundary conditions (2a) and(2c), we have a two-dimensional space of solutions. We find a basis for this space by looking for solution vectors of the form $\pi_{r,k} = \xi^k$; these must satisfy

$$p_r\pi_{r,k-1} + \pi_{r,k}(p_s - z) + \pi_{r,k+1}p_l = 0.$$

Therefore ξ must be a root of the *characteristic equation*

$$\xi^2 + \frac{(p_s - z)}{p_l}\xi + \frac{p_r}{p_l} = 0. \tag{3}$$

If the characteristic equation has two distinct roots, ξ_1 and, then the solution must take the form $\pi_{r,k} = c_1\xi_1^k + c_2\xi_2^k$. In this case, we can rewrite (3) as

$$\xi^2 + \frac{(p_s - z)}{p_l}\xi + \frac{p_r}{p_l} = (\xi - \xi_1)(\xi - \xi_2) = \xi^2 - (\xi_1 + \xi_2)\xi + \xi_1\xi_2.$$

This implies $\xi_1\xi_2 = \frac{p_r}{p_l}$ and $-\frac{(p_s - z)}{p_l} = (\xi_1 + \xi_2)$. Using this fact, we rewrite the boundary conditions in matrix form as:

$$\begin{bmatrix} p_s + p_l - z + \xi_1^1 p_l & p_s + p_l - z + \xi_2 p_l \\ (p_s + p_r - z)\xi_1^N + \xi_1^{N-1}p_r & (p_s + p_r - z)\xi_2^N + \xi_2^{N-1}p_r \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

A nontrivial solution to this equation exists if and only if the determinant of the matrix vanishes. Therefore

$$(\xi_1 - 1)\left(\xi_1 - \frac{p_r}{p_l}\right)\left(\xi_1^{2(N+1)} - \left(\frac{p_r}{p_l}\right)^{N+1}\right) = 0.$$

The case when we have the double root $\xi_1 = \xi_2 = \sqrt{p_r/p_l}$ does not correspond to a possible eigenvector for $p_r \neq p_l$. For $p_r = p_l$, the choice $\xi_1 = \xi_2 = 1$ corresponds to the eigenvector e^T. Thus, the eigenvectors in complex form for $0 \leq m < N+1$ are

$$\begin{aligned}
\pi_{0,k} &= c_1 + c_2\left(\frac{p_r}{p_l}\right)^k \\
\pi_{m,k} &= c_1\left(\sqrt{\left(\frac{p_r}{p_l}\right)} e^{\frac{\pi i m}{N+1}}\right)^k + c_2\left(\sqrt{\left(\frac{p_r}{p_l}\right)} e^{-\frac{\pi i m}{N+1}}\right)^k.
\end{aligned}$$

Note that $\frac{(z-p_s)}{p_l} = (\xi_1 + \xi_2) = 2\cos(\frac{m\pi}{N+1})$; therefore:

$$\begin{aligned}
z_m &= 2p_l\sqrt{\left(\frac{p_r}{p_l}\right)}\cos\left(\frac{m\pi}{N+1}\right) + p_s = 2\sqrt{p_r p_l}\cos\left(\frac{m\pi}{N+1}\right) + p_s \\
z_0 &= 1.
\end{aligned}$$

□

The second largest eigenvalue is $\lambda_2 = z_1 = 2\sqrt{p_r p_l}\cos(\frac{\pi}{N+1}) + p_s$, which increases monotonically toward $1 - (\sqrt{p_r} - \sqrt{p_l})^2$ as $N \to \infty$.

Theorem 3.4. *The stationary distribution of the bounded random walk has the form*

$$\pi_{0,k} \propto \left(\frac{p_r}{p_l}\right)^k.$$

Proof. For $z = 1$, $\pi_{0,k} = c_0\left(\frac{p_r}{p_l}\right)^k + c_1$, and by (2a),

$$\begin{aligned}
c_0\left(p_s + p_l - 1 + \frac{p_r}{p_l}p_l\right) &= -c_1(p_s + p_l - 1 + p_l) \\
c_1(p_r - p_l) &= 0.
\end{aligned}$$

Therefore $c_1 = 0$ for $p_r \neq p_l$. For the case $p_r = p_l$, we have $p_r/p_l = 1$ and the stationary distribution is $\pi_{0,k} = (N+1)^{-1} \propto 1$. □

While computing the eigenvectors explicitly is possible, it is much easier to use *reversibility* to compute a bound on the condition number.

Theorem 3.5. *The condition number for the bounded random walk is*

$$\frac{1}{N+1}\kappa_2(V) \leq \kappa_1(V) \leq (N+1)\kappa_2(V), \tag{4}$$

where

$$\kappa_2(V) = \left(\frac{p_l}{p_r}\right)^{N/2}. \tag{5}$$

Proof. By standard norm inequalities [2, §2.3.2], for any matrix V of size $(N+1)\times(N+1)$,

$$\frac{1}{\sqrt{N+1}}\|V\|_2 \le \|V\|_1 \le \|V\|_2\sqrt{N+1}.$$

Using the definitions $\kappa_1(V) = \|V\|_1\|V^{-1}\|_1$ and $\kappa_1(V) = \|V\|_2\|V^{-1}\|_2$ implies (4). To see (5), from Theorem 2.6 write

$$\begin{aligned}\kappa_2(V) &= \|V\|_2\|V^{-1}\|_2\\ &= \|\Pi^{\frac{-1}{2}}W\|_2\|W^T\Pi^{\frac{1}{2}}\|_2\\ &= \|\Pi^{\frac{-1}{2}}\|_2\|\Pi^{\frac{1}{2}}\|_2\\ &= \left(\frac{p_l}{p_r}\right)^{N/2}.\end{aligned}$$

□

Finally, we bound $\kappa_1(V)\lambda_2^k$ by

$$|\kappa_1(V)\lambda_2^k| \le N\left(\frac{p_l}{p_r}\right)^{-\frac{N}{2}}\left(2\sqrt{p_rp_l}\cos\left(\frac{\pi}{N+1}\right)+p_s\right)^k. \tag{6}$$

Combining (6) with Theorem 2.10 gives us (1).

4. Biased Random Walk

We now consider the special case of the *biased* random walk, in which $p_l > p_r$. In this case, the second largest eigenvalue is $\lambda_2 = z_1 = 2\sqrt{p_rp_l}\cos(\frac{\pi}{N+1})+p_s$, which increases monotonically toward $1-(\sqrt{p_r}-\sqrt{p_l})^2 < 1$ as $N\to\infty$. This upper bound on λ_2 gives an upper bound on the convergence:

$$\|\pi_k-\pi_*\|_1 \le |\kappa_1(V)\lambda_2^k| \le (N+1)\left(\frac{p_l}{p_r}\right)^{N/2}\left(1-(\sqrt{p_r}-\sqrt{p_l})^2\right)^k. \tag{7}$$

Using (7), we find that for any fixed ϵ, $\|\pi_k-\pi_*\|_1 < \epsilon$ whenever

$$k \ge \left|\frac{(\log(p_l)-\log(p_r))}{2\log\left(1-(\sqrt{p_r}-\sqrt{p_l})^2\right)}\right| N + O(\log(N)) = CN + O(\log(N)).$$

However, we expect (and will eventually prove) convergence happens in about $N/(p_l-p_r)$ steps (the approximate maximum number of steps for the mode of the unbounded distribution to reach the endpoint of the domain), so in many cases C appears to be large. For example, for $p_r = 0.2$ and $p_l = 0.3$, we have, $C \approx 20$ while $|p_r-p_l|^{-1} = 10$. Figure 4.1 shows the looseness of the naïve bound based on the condition number.

4.1. **The Unbounded Distribution for the Biased Random Walk.** Our intuition for why the biased random walk should converge in about $N/(p_l-p_r)$ steps is that this is the number of steps for the expected position of the walk to travel from N to 0 when there are no boundary conditions. Following our intuition, we expect a fruitful way to examine the biased random walk on $N+1$ points is to compare it to a biased random walk on an unbounded domain. The two random walks behave similarly as long as the walker on the bounded domain is not too likely to come close to the boundary. We devote this section to proving several properties of the biased random walk on an unbounded domain and relating them to the biased random walk.

Definition 4.1. At step k of the *random walk on unbounded domain*, we define the probability distribution ν_{k+1} by the equation:

$$\nu_{k+1,j} = p_r \nu_{k,j-1} + p_s \nu_{k,j} + p_l \nu_{k,j+1}.$$

Assume that the walk starts from the position x_0, i.e.

$$\nu_{0,j} = \begin{cases} 1, & j = x_0 \\ 0, & \text{otherwise.} \end{cases}$$

The distribution ν_k has mean $\mu_k = k(p_r - p_l) + p_0$ and standard deviation $\sigma_k = \sqrt{k(p_r + p_l - (p_r - p_l)^2)}$; via the central limit theorem, as k gets larger ν_k will rapidly approach a Gaussian distribution. The convergence of the distribution towards Gaussian makes ν_k easier to analyze than π_k. In particular, we will make use of the fact that we have an exponentially decaying bound on the tails of ν_k based on Chernoff's inequalities [5, p. 451].

Theorem 4.2. *Let ν_k be the distribution for the kth step of the unbounded random walk starting at x_0, and let $\mu = p_r - pl$. Then if $X_k \sim \nu_k$, for any $A > E[X] = x_0 + k\mu$,*

$$P\{X_k \geq A\} \equiv \sum_{j \geq A} \nu_k(j) \leq \exp\left(-\frac{(x_0 + k\mu - A)^2}{2k}\right). \tag{8}$$

Similarly, for any $A < E[X] = x_0 + k\mu$,

$$P\{X_k \leq A\} \equiv \sum_{j \leq A} \nu_k(j) \leq \exp\left(-\frac{(x_0 + k\mu - A)^2}{2k}\right). \tag{9}$$

Proof. The moment generating function for one step of the random walk is $M(t) = p_l e^{-t} + p_s + p_r e^t$. Now define $f(t) = \log M(t)$. By Taylor expansion $f(t) \leq \mu t + \frac{1}{2}t^2$, where we define $\mu = p_r - p_l$. Thus

$$M(t) = e^{f(t)} \leq \exp\left((p_r - p_l)t + \frac{1}{2}t^2\right).$$

If X_k represents the kth step of the biased random walk, then

$$M_{X_k}(t) = \exp^{tx_0} M(t)^j \leq \exp\left(\mu j t + \frac{j}{2}t^2\right).$$

According to Chernoff's inequality, if the moment generating function of X_k is $M_X(t) = E[e^{tX_k}]$, then for any $t > 0$, $P\{X_k \geq A\} \leq e^{-At} M_{X_k}(t)$. Therefore

$$P\{X_k \geq A\} \leq \exp\left(t(x_0 - N) + \mu j t + \frac{j}{2}t^2\right). \tag{10}$$

The right hand side of (10) is smallest when $t_* = -(x_0 + k\mu - A)/k$. Thus the optimum lower bound is (8). Similarly, when $A < E[X_k]$, $t_* < 0$ and the right-hand side of (10) bounds $P\{X_k \leq A\}$; this gives us (9). □

Therefore, when the expected value of the unbounded random walk is not too near the edges of the bounded domain, the unbounded walk has an exponentially small amount of probability mass outside the bounded domain. Our next focus will be to compare ν_k and π_k.

Theorem 4.3. *Let $\hat{P}$ be the transition matrix for the biased random walk on the unbounded domain ($\nu_k \hat{P} = \nu_{k+1}$) and let P be the transition matrix of the bounded domain extended to the infinite domain as follows*

$$P_{ij} = \begin{cases} p_s, & i = j \notin \{-1, 0, N, N+1\} \\ p_l, & i = j+1 \notin \{0, N+1\} \\ p_r, & i = j-1 \notin \{-1, N\} \\ p_s + p_l, & i = j \in \{0, N+1\} \\ p_s + p_r, & i = j \in \{-1, N\}. \end{cases}$$

If $\nu_0 = \pi_0$, then

$$\|\pi_k - \nu_k\| \leq \sum_{i=0}^{k-1} \left(\sum_{j \leq 0} \nu_k(j) + \sum_{N \leq j} \nu_k(j) \right).$$

Proof.

$$\begin{aligned} \pi_{k+1} - \nu_{k+1} &= (\pi_k)P + \nu_k(\hat{P}) = (\pi_k - \nu_k)P + \nu_k(P - \hat{P}) \\ \|\pi_{k+1} - \nu_{k+1}\|_1 &\leq \|(\pi_k - \nu_k)\|_1 + \|\nu_k(P - \hat{P})\|_1 \\ \|\pi_k - \nu_k\| &\leq \|(\pi_0 - \nu_0)\|_1 + \sum_{i=0}^{k-1} \|\nu_i(P - \hat{P})\|_1. \end{aligned}$$

Apply $\nu_0 = \pi_0$ to make the first term vanish, leaving

$$\|\pi_k - \nu_k\|_1 \leq \sum_{i=0}^{k-1} \|\nu_i(P - \hat{P})\|_1;$$

and because $P([1, N-1]) = \hat{P}([1, N-1])$,

$$\|\nu_k(P - \hat{P})\|_1 \leq \sum_{j \leq 0} \nu_k(j) + \sum_{N \leq j} \nu_k(j).$$

□

While one of the tail sums will eventually grow without bound ($\sum_{i=0}^{k} \sum_{j \leq 0} \nu_i(j)$), the other tail sum remains bounded:

Theorem 4.4. *Let ν_k be the kth step of the unbounded distribution starting at x_0 with $\mu = p_r - pl$. Then*

$$R_k := \sum_{i=0}^{k} \sum_{N \leq j} \nu_i(j) < \frac{\exp\left(-(N - x_0)\left(\frac{\mu^2}{2} - \mu\right)\right)}{1 - \exp(-\frac{\mu^2}{2})}.$$

Proof. Using (8),

$$\begin{aligned} R_k &= \sum_{i=0}^{\infty}\sum_{N\le j}\nu_i(j) = \sum_{i=N-x_0}^{\infty}\sum_{N\le j}\nu_i(j) \\ &\le \sum_{j=N-x_0}^{\infty}\exp\left(-\left((x_0-N)\mu+\frac{j\mu^2}{2}\right)\right) \\ &= \sum_{j=0}^{\infty}\exp\left(-\left((x_0-N)\mu+\frac{(N-x_0)\mu^2}{2}+\frac{j\mu^2}{2}\right)\right) \\ &= \frac{\exp\left(-(N-x_0)\left(\frac{\mu^2}{2}-\mu\right)\right)}{1-\exp\left(-\frac{\mu^2}{2}\right)}. \end{aligned}$$

□

Theorem 4.4 implies that $R_k < \epsilon$ for all k provided $N - x_0 \ge \beta$, where

$$\beta := \frac{\log\epsilon + \log\left(1-\exp(\frac{-\mu^2}{2})\right)}{\mu-\frac{\mu^2}{2}}. \tag{11}$$

Thus the number of steps from the right boundary needed to bound the right tail sum R_k is *independent* of N.

Theorem 4.5. *Let ν_k be the kth step of the unbounded distribution starting at x_0 with $\mu = p_r - pl$. Then for any $\gamma > 0$ and any $k < -(x_0 - \gamma\sqrt{x_0})/\mu$,*

$$L_k := \sum_{i=0}^{k-1}\sum_{j<0}\nu_i(j) \le \frac{2\sqrt{x_0}}{\gamma\mu^2}\exp\left(\frac{\mu\gamma^2}{2}\right). \tag{12}$$

Proof. Using the Chernoff bound (9),

$$L_k = \sum_{i=0}^{k-1}\sum_{j<0}\nu_i(j) \le \sum_{j=0}^{k-1}\exp\left(-\frac{(x_0)^2+2x_0j\mu+(j\mu)^2}{2j}\right).$$

If we choose $k = -\frac{x_0}{\mu}(1-\delta)$ for some arbitrary δ,

$$\begin{aligned} L_k &\le \sum_{j}^{k-1}\exp\left(-\frac{(x_0)^2+2x_0j\mu+(j\mu)^2}{2j}\right) \\ &\le \sum_{j}^{k-1}\exp\left(-\frac{(x_0+k\mu)}{2k}(x_0+j\mu)\right) = \sum_{j}^{k-1}\exp\left(\frac{\delta\mu}{2(1-\delta)}(x_0+j\mu)\right) \\ &\le \sum_{j}^{k-1}\exp\left(\frac{\delta\mu}{2}(x_0+j\mu)\right) \le \int_{-\infty}^{k}\exp\left(\frac{\delta\mu}{2}(x_0+t\mu)\right)dt \\ &= \frac{2}{\delta\mu^2}\exp\left(\frac{\delta\mu}{2}(x_0+k\mu)\right) = \frac{2}{\delta\mu^2}\exp\left(\frac{\delta^2\mu x_0}{2}\right). \end{aligned}$$

Choose $\delta = \frac{\gamma}{\sqrt{x_0}}$ which corresponds to

$$k = -\frac{x_0}{\mu} + \gamma\frac{\sqrt{x_0}}{\mu}.$$

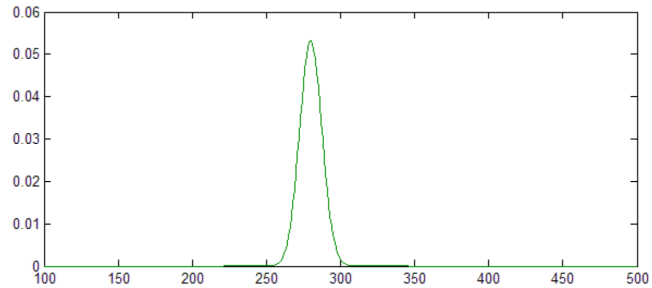

FIGURE 1. The blue line depicts ν_k and the green line is π_k on the bounded set (200,400). In this example, the starting position was chosen in the middle of the bounded interval and was only run for 100 iterations. Because no probability mass of ν_k lies outside the boundary, our theorems predict that $\|\nu_k - \pi_k\|_1 = 0$. As seen, they are identical.

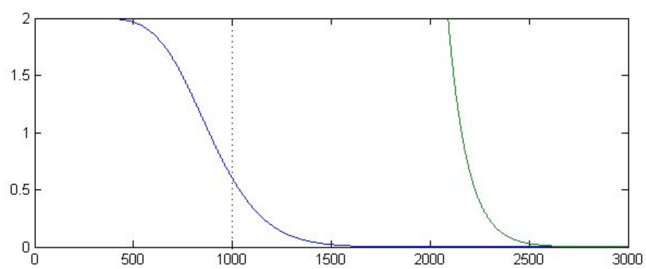

FIGURE 2. The blue line represents $\|\pi_k - \pi_*\|$ and the green line represents the condition number bound. Notice that the naïve condition number bound is very loose.

Therefore

$$L_k \leq \frac{2\sqrt{x_0}}{\gamma\mu^2} \exp\left(\frac{\mu\gamma^2}{2}\right)$$

is the final result. □

For any $\epsilon > 0$, therefore, we have $L_k < \epsilon$ for any γ that satisfies

$$\gamma^2 - \frac{2}{\mu}\log(\gamma) > -\frac{2}{\mu}\left(\log\left(\frac{2}{\epsilon\mu^2}\right) + \frac{1}{2}\log x_0\right). \tag{13}$$

The condition (13) is satisfied whenever γ is bounded from below by

$$\gamma_* = \max\left(1, \sqrt{-\frac{2}{\mu}\left(\log\left(\frac{2}{\epsilon\mu^2}\right) + \frac{1}{2}\log x_0\right)}\right) = O\left(\sqrt{\log(x_0)}\right).$$

Therefore, the left tail sum L_k will be less than ϵ for $k < -x_0/\mu - O(\sqrt{x_0 \log x_0})$.

Figure 4.1 shows the comparison between the distributions for the bounded and unbounded biased random walks far from the boundary. The two distributions are visually indistinguishable.

4.2. **Application of Bounds to the Biased Random Walk.** Our goal in this section is to combine the insights gained in Section 4.1 about the initial similarity of the bounded and unbounded biased random walks with our spectral bounds from Section 3. In order to do this, we will need a slightly tighter variant of the spectral bound that uses information about the support of the initial distribution.

Lemma 4.6. *Let P be the transition matrix for the biased random walk on $N+1$ points. Suppose π_0 has support only on $0, \ldots, L$. Then*

$$\|(\pi_0 - \pi^*)P^k\|_1 \leq \sqrt{N+1}\left(\frac{p_l}{p_r}\right)^{L/2} |\lambda_2|^k.$$

Proof. As before, write the eigendecomposition $P = \Pi^{-1/2}W\Lambda W^T\Pi^{1/2}$. Recall that $(\pi_0 - \pi^*)P^k = \pi_0 A^k$ where $A = \Pi^{-1/2}W\hat{\Lambda}W^T\Pi^{1/2}$, with $\hat{\Lambda}$ the same as Λ, but with the eigenvalue at one replaced with zero.

Now compute

$$\|\pi_0 A^k\|_1 \leq \sqrt{N+1}\|\pi_0 A^k\|_2 = \sqrt{N+1}\|\pi_0\Pi^{-1/2}W\hat{\Lambda}^k W^T\Pi^{1/2}\|_2 \tag{14}$$

$$\leq \sqrt{N+1}\|\pi_0\Pi^{-1/2}\|_2\|W\|_2\|\hat{\Lambda}\|_2^k\|W^T\|_2\|\Pi^{1/2}\|_2 \tag{15}$$

$$= \sqrt{N+1}\|\pi_0\Pi^{-1/2}\|_2\left(\max_j \sqrt{\pi^*(j)}\right)|\lambda_2|^k \tag{16}$$

Let $\alpha = p_r/p_l$, and let $C = (1-\alpha)/(1-\alpha^{N+1})$ so that $\pi^*(j) = C\alpha^j$ and $\max_j \pi^*(j) = C$. Then

$$\left(\pi_0\Pi^{-1/2}\right)_j^2 = \frac{(\pi_0)_j^2}{C\alpha^j} \leq \frac{(\pi_0)_j}{C\alpha^L}.$$

Therefore

$$\|\pi_0\Pi^{-1/2}\|_2^2 = \sum_{j=0}^{N}\left(\pi_0\Pi^{-1/2}\right)_j^2 \leq \frac{1}{C\alpha^L}. \tag{17}$$

Combining (17) with (16) yields Lemma 4.6's result. □

Of course, in practice we never get to the point where the tail is exactly zero between some interesting L and N. Therefore, we need the following refinement of Lemma 4.6.

Lemma 4.7. *Let $E = \{j : L < j \leq N\}$ for L arbitrarily chosen. Then*

$$\|(\pi_0 - \pi^*)P^k\|_1 \leq \sqrt{N+1}\left(\frac{p_l}{p_r}\right)^{L/2} |\lambda_2|^k + 2\pi_0(E)$$

Proof. Construct $\widetilde{\pi}$ by shifting all the probability in $\pi_0(E)$ onto the first index; that is,

$$\widetilde{\pi}(j) = \begin{cases} \pi_0(j) + \pi_0(E), & j = 0 \\ \pi_0(j), & 0 < j \leq L \\ 0, & L < j \leq N. \end{cases}$$

By the triangle inequality

$$\|(\pi_0 - \pi^*)P^k\|_1 \leq \|(\pi_0 - \tilde{\pi})P^k\|_1 + \|(\widetilde{\pi} - \pi^*)P^k\|_1. \tag{18}$$

To bound the first term in (18), note that $\|P\|_1 = 1$, so

$$\|(\pi_0 - \tilde{\pi})P^k\|_1 \leq \|\pi_0 - \tilde{\pi}\|_1 = 2\pi_0(E)).$$

Bound the second term using Lemma 4.6 to complete the proof. □

From here, our argument is as follows. For about $N/|\mu|$ steps, the bounded biased random walk behaves like the unbounded random walk. This means that most of the probability is concentrated in an interval around the mean; and that interval is disjoint from an interval containing most of the probability for the stationary distribution. Therefore, the distribution remains far from stationarity for about $N/|\mu|$ steps. When the probability mass reaches the endpoints, it is concentrated in a region of size $O(\sqrt{N})$, and Lemma 4.7 tells us that convergence occurs in an additional $O(\sqrt{N})$ steps.

Theorem 4.8. *Assume x_0 is chosen so that $N - x_0 > \beta(\epsilon)$ as defined in* (11). *Then for $k < x_0/|\mu| - O(\sqrt{x_0 \log x_0})$,*

$$\|\pi_k - \pi^*\|_1 > 2 - 6\epsilon, \tag{19}$$

and for $k > x_0/|\mu| + O(\sqrt{x_0 \log x_0})$,

$$\|\pi_k - \pi^*\|_1 < 7\epsilon. \tag{20}$$

Proof. From Theorems 4.3, 4.4, and 4.5, we know $\|\pi_k - \nu_k\| \leq 2\epsilon$ for $k < x_0/|\mu| - \gamma\sqrt{x_0}$ for $\gamma = O(\sqrt{\log x_0})$. By the triangle inequality, when k is so restricted,

$$\|\pi_k - \pi^*\|_1 > \|\nu_k - \pi^*\| - 2\epsilon.$$

Note that if E is any set,

$$\begin{aligned} \|\nu_k - \pi^*\|_1 &= \sum_{j \in E} |\nu_k(j) - \pi^*(j)| + \sum_{j \notin E} |\nu_k(j) - \pi^*(j)| \\ &> -\nu_k(E) + \pi^*(E) - \pi^*(E^c) + \nu_k(E^c) \\ &= -\nu_k(E) + (1 - \pi^*(E^c)) - \pi^*(E^c) + (1 - \nu_k(E)) \\ &= 2(1 - \nu_k(E) - \pi^*(E^c)). \end{aligned}$$

We will apply this bound with $E = \{0, \ldots, L\}$.

If we choose $L > \log(\epsilon)/\log(p_r/p_l)$, then we know that $\pi_k\{L+1, \ldots, N\} \leq \epsilon$. By the Chernoff inequality (9), we know that $\nu_k\{0, \ldots, L\} < \epsilon$ if

$$\exp\left(-\frac{(x_0 + k\mu - L)^2}{2k}\right) < \epsilon,$$

which holds provided

$$k < \frac{x_0 - L}{|\mu|} + \frac{\log(\epsilon^{-1})}{\mu^2}. \tag{21}$$

Therefore if k satisfies (21), we have

$$\|\nu_k - \pi_*\|_1 \geq 2(1 - 2\epsilon).$$

This concludes the proof of (19).

We now turn to the proof of (20). By Lemma 4.7, for any k_* and for any $0 < M < N$,

$$\|\pi_{k_*+l} - \pi^*\|_1 \leq \sqrt{N+1}\left(\frac{p_l}{p_r}\right)^{M/2} |\lambda_2|^l + 2\pi_{k_*}\{M+1, \ldots, N\}. \tag{22}$$

Now define

$$k_* = \min\left(\frac{x_0 - L}{|\mu|} + \frac{\log(\epsilon^{-1})}{\mu^2}, \frac{x_0}{|\mu|} - \gamma_* \sqrt{x_0}\right) = \frac{x_0}{|\mu|} + O(\sqrt{x_0 \log x_0}).$$

We know that

$$\|\nu_{k_*} - \pi_{k_*}\|_1 < 2\epsilon.$$

Therefore, for any M,

$$2\pi_{k_*}\{M+1, \ldots, N\} \leq 2\left(\nu_{k_*}\{M+1, \ldots, N\} + \epsilon\right).$$

Now choose M using (8) so that

$$\nu_{k_*}\{M+1, \ldots, N\} < \exp\left(-\frac{(x_0 + k_*\mu - M)^2}{2k_*}\right) = \epsilon,$$

that is,

$$M \geq x_0 + k_*\mu - \sqrt{2k_* \log(\epsilon^{-1})} = O(\sqrt{x_0 \log x_0}).$$

Therefore,

$$\sqrt{N+1}\left(\frac{p_l}{p_r}\right)^{M/2} |\lambda_2|^l = \epsilon$$

for $l = O(\sqrt{x_0 \log x_0})$. In this case, (22) gives us

$$\|\pi_{k_*+l} - \pi^*\|_1 \leq 7\epsilon,$$

and $k_* + l = x_0/|\mu| + O(\sqrt{x_0 \log x_0})$. □

According to Theorem 4.8, $\|\pi_k - \pi_*\|$ goes from $2 - O(\epsilon)$ to $O(\epsilon)$ in $\max_{1 \leq x_0 \leq N} O(\sqrt{x_0 \log x_0}) = O(\sqrt{N \log N})$ steps around $k = N/|\mu|$. Therefore, the biased random walk has a *cutoff*:

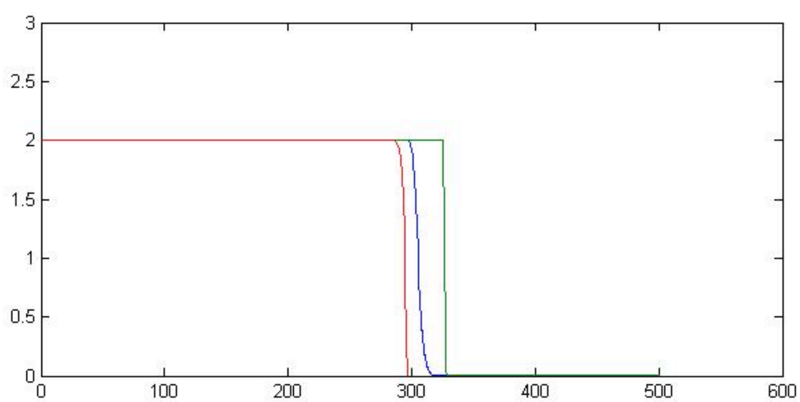

FIGURE 3. This picture is the biased random walk with $|p_r - p_l| \approx 1$ and N=300. The comparison to Figure 2 illustrates the cutoff. The green and red lines depict upper and lower bounds proven in Theorem 4.7 and Theorem 4.8. Compared to the upper bound in Figure 2, there is an improvement in the upper bound's tightness.

Definition 4.9. For a discrete Markov chain on N states, let a_n, b_n be functions tending towards infinity, with $\frac{b_n}{a_n}$ approaching zero, i.e., $b_n = o(a_n)$. A Markov chain satisfies a a_n, b_n cutoff if for some starting state x_n and all fixed real θ, with $k_0 = \lfloor a_n + b_n\theta \rfloor$,

$$\|\pi_{k_n} - \pi^*\|_1 \to c(\theta),$$

where $c(\theta)$ is a function that tends toward zero for θ tending toward infinity and θ tending toward 2 as θ tends toward minus infinity [1].

In particular, the biased random walk has a $N/|\mu|$, $\sqrt{N \log N}$ cutoff.

5. Unbiased Random Walk

Our next example is the *unbiased random walk*: $p_l = p_r = p$. The unbiased random walk is the limiting case of the biased random walk as we let $p_l - p_r \to 0$. There are many similarities between the biased and unbiased cases, but a few interesting things change in the unbiased case. In particular, the unbiased walk does not have a cutoff, but converges in $k \sim O(N^2)$ steps.

In the case of the biased random walk, the eigenvalues ranged from $1 - (\sqrt{p_r} + \sqrt{p_l})^2 \le z_m \le 1 - (\sqrt{p_r} - \sqrt{p_l})^2$ with a lower bound on $1 - |\lambda_2|$ of at least $|z_0 - z_1| > (\sqrt{p_r} - \sqrt{p_l})^2$. With the unbiased random walk, $-\frac{1}{3} < 1 - 4p < z_m < 1$. There is no minimum spectral gap.

The condition number bound in Theorem 2.10 is $\|\pi_k - \pi^*\|_1 \le (N+1)(1 - 2p + 2p\cos\frac{\pi}{N+1})^k$. Compared with the BRW, the condition number of the NRW bound grows modestly with increased N, but the second largest eigenvalue approaches one. Therefore, we will need to do a little more work to bound the convergence.

Theorem 5.1.

$$\|\pi^* - \pi_k\|_1 < 2(N+1)e^{-\frac{p\pi^2 k}{(N+1)^2}} + O(N^{-3})$$

Proof. From Theorem 2.10 we know $\|\pi^* - \pi_k\|_1 < |\lambda_2^k| \kappa_1(V)$. By (4), $\kappa_1(V) \le N+1$. We must show is that $\lambda_2 = (1 - 2p + 2p\cos\left(\frac{\pi}{N+1}\right))^k = e^{-\frac{p\pi^2}{(N+1)^2}} + O\left((N+1)^{-2}\right)$. Following [1], apply a Taylor expansion to $\cos\left(\frac{\pi}{N+1}\right)$;

$$\begin{aligned} 1 - 2p + 2p\cos\left(\frac{\pi}{N+1}\right) &= 1 - p\frac{\pi^2}{(N+1)^2} + O((N+1)^{-4}) \\ &= e^{-\frac{p\pi^2}{(N+1)^2}} + O((N+1)^{-4}). \end{aligned}$$

Therefore

$$\left(1 - 2p + 2p\cos\left(\frac{\pi}{N+1}\right)\right)^k = e^{-\frac{p\pi^2 k}{(N+1)^2}} + O((N+1)^{-4}).$$

□

Unlike the biased random walk, the unbiased random walk converges slowly at $O(N^2)$ and it does not have a cutoff.

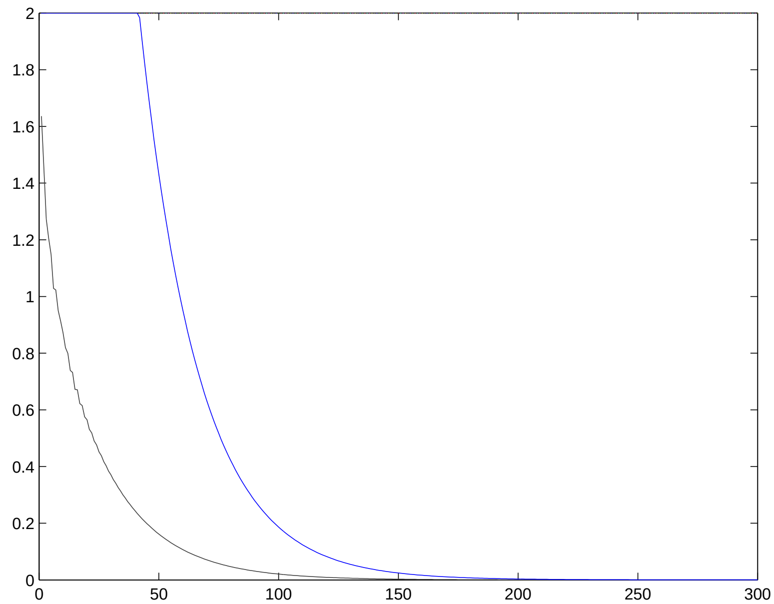

FIGURE 4. The dark blue line represents $\|\pi_k - \pi^*\|$ for the unbiased walk, while the light blue line is the upper bound. The picture shows the exponential decay of both $\|\pi_k - \pi^*\|$ and the upper bound, and illustrates the lack of a cutoff.

6. Conclusion

We have analyzed the dynamics of two Markov chains, the unbiased random walk and the biased random walk. We picked the biased random walk because our intuition tells us that the distribution during the transient period should look very different from the stationary distribution, an intuition which we have made precise.

For the case of the biased random walk on a large domain, the distribution will look Gaussian up until the point where the mode of the distribution hits the opposing boundary ($N/|\mu|$ steps). At this point, the shape will shift from Gaussian to exponential. An important tool in our analysis was the comparison of the biased random walk to a similar random walk on an unbounded domain. The comparison of the bounded and unbounded biased random walks let us show that for many steps the bounded walk is far from stationarity, and thus that there is a cutoff.

We have also shown that the convergence of the unbiased random walk is qualitatively different from the convergence of the biased walk. In the unbiased case, the initial probability distribution diffuses, approaching a uniform stationary distribution. The convergence is slow, but steady, and for any fixed ϵ it takes $O(N^2)$ steps before $\|\pi_k - \pi^*\| < \epsilon$. There is no cutoff.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. David Bindel, for his help on this project. I am extremely lucky to have his help. He was instrumental in assisting me in many ways, like showing me the ropes in MATLAB, pushing me through even the toughest proofs, and encouraging me through the difficult challenges. His help was invaluable to me throughout the entire summer.

Next, I would like to thank the Courant Institute for their part. Without the financial assistance, or the help of people like Jackie Klein and Fred Greenleaf, this would not have been possible.

References

[1] Persi Daiconis, *The cutoff phenomenon in finite Markov chains*, Proc. Natl. Acad. Sci. **93** (1996), 1659–1664.

[2] Gene H. Golub and Charles F. Van Loan, *Matrix computations*, The John Hopkins University Press, 1989.

[3] Peter D. Lax, *Linear algebra*, Wiley Interscience, 1997.

[4] Carl D. Meyer, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Review **31** (1989), no. 2, 240–272.

[5] Sheldon Ross, *Introduction to Probability Theory*, Prentice Hall, 2006.

[6] Laurent Saloff-Coste, *Random walks on finite groups*, Probability on discrete structures (Harry Kesten, ed.), Encyclopaedia Math. Sci., vol. 110, Springer, 2004, pp. 263–364.

[7] Lloyd N. Trefethen and Mark Embree, *Spectra and pseudospectra*, Princeton University Press, 2005.