Bounding the mixing time via the spectral gap
Applications: random walk on cycle and hypercube
Infinite networks

Review

Modern Discrete Probability

VI - Spectral Techniques

Background

Sébastien Roch

UW–Madison
Mathematics

December 22, 2014
1. Review
2. Bounding the mixing time via the spectral gap
3. Applications: random walk on cycle and hypercube
4. Infinite networks
Theorem (Convergence to stationarity)

Consider a finite state space $V$. Suppose the transition matrix $P$ is irreducible, aperiodic and has stationary distribution $\pi$. Then, for all $x, y$, $P^t(x, y) \to \pi(y)$ as $t \to +\infty$.

For probability measures $\mu, \nu$ on $V$, let their total variation distance be $\|\mu - \nu\|_{TV} := \sup_{A \subseteq V} |\mu(A) - \nu(A)|$.

Definition (Mixing time)

The mixing time is

$$t_{mix}(\varepsilon) := \min\{t \geq 0 : d(t) \leq \varepsilon\},$$

where $d(t) := \max_{x \in V} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV}$.
Definition (Separation distance)

The *separation distance* is defined as

\[ s_x(t) := \max_{y \in V} \left[ 1 - \frac{P^t(x, y)}{\pi(y)} \right], \]

and we let \( s(t) := \max_{x \in V} s_x(t). \)

Because both \( \{\pi(y)\} \) and \( \{P^t(x, y)\} \) are non-negative and sum to 1, we have that \( s_x(t) \geq 0. \)

Lemma (Separation distance v. total variation distance)

\[ d(t) \leq s(t). \]
Proof: Because $1 = \sum_y \pi(y) = \sum_y P^t(x,y)$,

$$\sum_{y : P^t(x,y) < \pi(y)} \left[ \pi(y) - P^t(x,y) \right] = \sum_{y : P^t(x,y) \geq \pi(y)} \left[ P^t(x,y) - \pi(y) \right].$$

So

$$\| P^t(x, \cdot) - \pi(\cdot) \|_{TV} = \frac{1}{2} \sum_y \left| \pi(y) - P^t(x,y) \right| = \sum_{y : P^t(x,y) < \pi(y)} \left[ \pi(y) - P^t(x,y) \right] = \sum_{y : P^t(x,y) < \pi(y)} \pi(y) \left[ 1 - \frac{P^t(x,y)}{\pi(y)} \right] \leq S_x(t).$$
Reversible chains

Definition (Reversible chain)

A transition matrix $P$ is *reversible* w.r.t. a measure $\eta$ if
$\eta(x)P(x, y) = \eta(y)P(y, x)$ for all $x, y \in V$. By summing over $y$, such a measure is necessarily stationary.
Recall:

**Definition (Random walk on a graph)**

Let $G = (V, E)$ be a finite or countable, locally finite graph. A **simple random walk** on $G$ is the Markov chain on $V$, started at an arbitrary vertex, which at each time picks a uniformly chosen neighbor of the current state.

Let $(X_t)$ be simple random walk on a connected graph $G$. Then $(X_t)$ is reversible w.r.t. $\eta(v) := \delta(v)$, where $\delta(v)$ is the degree of vertex $v$. 

Example I
Definition (Random walk on a network)

Let $G = (V, E)$ be a finite or countable, locally finite graph. Let $c : E \to \mathbb{R}^+$ be a positive edge weight function on $G$. We call $\mathcal{N} = (G, c)$ a network. Random walk on $\mathcal{N}$ is the Markov chain on $V$, started at an arbitrary vertex, which at each time picks a neighbor of the current state proportionally to the weight of the corresponding edge.

Any countable, reversible Markov chain can be seen as a random walk on a network (not necessarily locally finite) by setting $c(e) := \pi(x)P(x, y) = \pi(y)P(y, x)$ for all $e = \{x, y\} \in E$. Let $(X_t)$ be a random walk on a network $\mathcal{N} = (G, c)$. Then $(X_t)$ is reversible w.r.t. $\eta(v) := c(v)$, where $c(v) := \sum_{x \sim v} c(v, x)$. 

Sébastien Roch, UW–Madison

Modern Discrete Probability – Spectral Techniques
We let $n := |V| < +\infty$. Assume that $P$ is irreducible and reversible w.r.t. its stationary distribution $\pi > 0$. Define

$$\langle f, g \rangle_\pi := \sum_{x \in V} \pi(x)f(x)g(x), \quad \|f\|_\pi^2 := \langle f, f \rangle_\pi,$$

$$(Pf)(x) := \sum_y P(x, y)f(y).$$

We let $\ell^2(V, \pi)$ be the Hilbert space of real-valued functions on $V$ equipped with the inner product $\langle \cdot, \cdot \rangle_\pi$ (equivalent to the vector space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_\pi)$).

**Theorem**

*There is an orthonormal basis of $\ell^2(V, \pi)$ formed of eigenfunctions $\{f_j\}_{j=1}^n$ of $P$ with real eigenvalues $\{\lambda_j\}_{j=1}^n$. We can take $f_1 \equiv 1$ and $\lambda_1 = 1$.***


**Proof:** We work over \((\mathbb{R}^n, \langle \cdot, \cdot \rangle_\pi)\). Let \(D_\pi\) be the diagonal matrix with \(\pi\) on the diagonal. By reversibility,

\[
M(x, y) := \sqrt{\frac{\pi(x)}{\pi(y)}} P(x, y) = \sqrt{\frac{\pi(y)}{\pi(x)}} P(y, x) =: M(y, x).
\]

So \(M = (M(x, y))_{x,y} = D_\pi^{1/2} P D_\pi^{-1/2}\), as a symmetric matrix, has real eigenvectors \(\{\phi_j\}_{j=1}^n\) forming an orthonormal basis of \(\mathbb{R}^n\) with corresponding real eigenvalues \(\{\lambda_j\}_{j=1}^n\). Define \(f_j := D_\pi^{-1/2} \phi_j\). Then

\[
Pf_j = P D_\pi^{-1/2} \phi_j = D_\pi^{-1/2} D_\pi^{1/2} P D_\pi^{-1/2} \phi_j = D_\pi^{-1/2} M \phi_j = \lambda_j D_\pi^{-1/2} \phi_j = \lambda_j f_j,
\]
and

\[
\langle f_i, f_j \rangle_\pi = \langle D_\pi^{-1/2} \phi_i, D_\pi^{-1/2} \phi_j \rangle_\pi = \sum_x \pi(x) [\pi(x)^{-1/2} \phi_i(x)] [\pi(x)^{-1/2} \phi_j(x)] = \langle \phi_i, \phi_j \rangle.
\]

Because \(P\) is stochastic, the all-one vector is a right eigenvector of \(P\) with eigenvalue 1.
Lemma

For all \( j \neq 1 \), \( \sum_x \pi(x) f_j(x) = 0. \)

Proof: By orthonormality, \( \langle f_1, f_j \rangle_\pi = 0. \) Now use the fact that \( f_1 \equiv 1. \)

Let \( \delta_x(y) := \mathbb{1}_{\{x=y\}}. \)

Lemma

For all \( x, y \), \( \sum_{j=1}^n f_j(x)f_j(y) = \pi(x)^{-1} \delta_x(y). \)

Proof: Using the notation of the theorem, the matrix \( \Phi \) whose columns are the \( \phi_j \)s is unitary so \( \Phi \Phi' = I. \) That is, \( \sum_{j=1}^n \phi_j(x)\phi_j(y) = \delta_x(y), \) or \( \sum_{j=1}^n \sqrt{\pi(x)\pi(y)} f_j(x)f_j(y) = \delta_x(y). \) Rearranging gives the result.
Lemma

Let $g \in \ell^2(\mathcal{V}, \pi)$. Then $g = \sum_{j=1}^{n} \langle g, f_j \rangle \pi f_j$.

Proof: By the previous lemma, for all $x$

$$
\sum_{j=1}^{n} \langle g, f_j \rangle \pi f_j(x) = \sum_{j=1}^{n} \sum_{i} \pi(y) g(y) f_j(y) f_i(x) = \sum_{i} \pi(y) g(y) [\pi(x)^{-1} \delta_x(y)] = g(x).
$$

Lemma

Let $g \in \ell^2(\mathcal{V}, \pi)$. Then $\|g\|_\pi^2 = \sum_{j=1}^{n} \langle g, f_j \rangle_\pi^2$.

Proof: By the previous lemma,

$$
\|g\|_\pi^2 = \left\| \sum_{j=1}^{n} \langle g, f_j \rangle_\pi f_j \right\|_\pi^2 = \left\langle \sum_{i=1}^{n} \langle g, f_i \rangle_\pi f_i, \sum_{j=1}^{n} \langle g, f_j \rangle_\pi f_j \right\rangle_\pi = \sum_{i,j=1}^{n} \langle g, f_i \rangle_\pi \langle g, f_j \rangle_\pi \langle f_i, f_j \rangle_\pi,
$$
Let $P$ be finite, irreducible and reversible.

**Lemma**

Any eigenvalue $\lambda$ of $P$ satisfies $|\lambda| \leq 1$.

*Proof:* $Pf = \lambda f \implies |\lambda|\|f\|_\infty = \|Pf\|_\infty = \max_x |\sum_y P(x, y)f(y)| \leq \|f\|_\infty$.

We order the eigenvalues $1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq -1$. In fact:

**Lemma**

We have $\lambda_2 < 1$.

*Proof:* Any eigenfunction with eigenvalue 1 is $P$-harmonic. By Corollary 3.22 for a finite, irreducible chain the only harmonic functions are the constant functions. So the eigenspace corresponding to 1 is one-dimensional. Since all eigenvalues are real, we must have $\lambda_2 < 1$. 

\[\square\]
Theorem (Rayleigh’s quotient)

Let $P$ be finite, irreducible and reversible with respect to $\pi$. The second largest eigenvalue is characterized by

$$\lambda_2 = \sup \left\{ \frac{\langle f, Pf \rangle_\pi}{\langle f, f \rangle_\pi} : f \in \ell^2(V, \pi), \sum_x \pi(x)f(x) = 0 \right\}.$$

(Similarly, $\lambda_1 = \sup_{f \in \ell^2(V, \pi)} \frac{\langle f, Pf \rangle_\pi}{\langle f, f \rangle_\pi}$.)

Proof: Recalling that $f_1 \equiv 1$, the condition $\sum_x \pi(x)f(x) = 0$ is equivalent to $\langle f_1, f \rangle_\pi = 0$. For such an $f$, the eigendecomposition is

$$f = \sum_{j=1}^n \langle f, f_j \rangle_\pi f_j = \sum_{j=2}^n \langle f, f_j \rangle_\pi f_j,$$
and

\[ Pf = \sum_{j=2}^{n} \langle f, f_j \rangle_\pi \lambda_j f_j, \]

so that

\[
\frac{\langle f, Pf \rangle_\pi}{\langle f, f \rangle_\pi} = \frac{\sum_{i=2}^{n} \sum_{j=2}^{n} \langle f, f_i \rangle_\pi \langle f, f_j \rangle_\pi \lambda_j \langle f_i, f_j \rangle_\pi}{\sum_{j=2}^{n} \langle f, f_j \rangle_\pi^2} = \frac{\sum_{j=2}^{n} \langle f, f_j \rangle_\pi^2 \lambda_j}{\sum_{j=2}^{n} \langle f, f_j \rangle_\pi^2} \leq \lambda_2.
\]

Taking \( f = f_2 \) achieves the supremum.
The Dirichlet form is defined as $\mathcal{E}(f, g) := \langle f, (I - P)g \rangle_\pi$. Note that

\[
2\langle f, (I - P)f \rangle_\pi
= 2\langle f, f \rangle_\pi - 2\langle f, Pf \rangle_\pi
= \sum_x \pi(x) f(x)^2 + \sum_y \pi(y) f(y)^2 - 2 \sum_x \pi(x) f(x) f(y) P(x, y)
= \sum_{x,y} f(x)^2 \pi(x) P(x, y) + \sum_{x,y} f(y)^2 \pi(y) P(y, x) - 2 \sum_x \pi(x) f(x) f(y) P(x, y)
= \sum_{x,y} \pi(x) P(x, y) [f(x) - f(y)]^2 = 2\mathcal{E}(f)
\]

where

\[
\mathcal{E}(f) := \frac{1}{2} \sum_{x,y} c(x, y) [f(x) - f(y)]^2,
\]

is the Dirichlet energy encountered previously.
We note further that if \( \sum_x \pi(x)f(x) = 0 \) then

\[
\langle f, f \rangle_\pi = \langle f - \langle 1, f \rangle_\pi, f - \langle 1, f \rangle_\pi \rangle_\pi = \text{Var}_\pi[f],
\]

where the last expression denotes the variance under \( \pi \). So the variational characterization of \( \lambda_2 \) translates into

\[
\text{Var}_\pi[f] \leq \gamma^{-1} \mathcal{E}(f),
\]

for all \( f \) such that \( \sum_x \pi(x)f(x) = 0 \) (in fact for any \( f \) by considering \( f - \langle 1, f \rangle_\pi \) and noticing that both sides are unaffected by adding a constant), which is known as a Poincaré inequality.
Review

Bounding the mixing time via the spectral gap

Applications: random walk on cycle and hypercube

Infinite networks
Theorem

Let \( \{ f_j \}_{j=1}^n \) be the eigenfunctions of a reversible and irreducible transition matrix \( P \) with corresponding eigenvalues \( \{ \lambda_j \}_{j=1}^n \), as defined previously. Assume \( \lambda_1 \geq \cdots \geq \lambda_n \). We have the decomposition

\[
\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{j=2}^n f_j(x) f_j(y) \lambda_j^t.
\]
**Proof:** Let $F$ be the matrix whose columns are the eigenvectors $\{f_j\}_{j=1}^n$ and let $D_\lambda$ be the diagonal matrix with $\{\lambda_j\}_{j=1}^n$ on the diagonal. Using the notation of the eigenbasis theorem,

$$D_\pi^{1/2} P^t D_\pi^{-1/2} = M^t = (D_\pi^{1/2} F) D_\lambda^t (D_\pi^{1/2} F)' ,$$

which after rearranging becomes

$$P^t D_\pi^{-1} = F D_\lambda^t F'.$$
Example: two-state chain I

Let $V := \{0, 1\}$ and, for $\alpha, \beta \in (0, 1)$,

$$P := \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

Observe that $P$ is reversible w.r.t. to the stationary distribution

$$\pi := \left( \frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

We know that $f_1 \equiv 1$ is an eigenfunction with eigenvalue 1. As can be checked by direct computation, the other eigenfunction (in vector form) is

$$f_2 := \left( \sqrt{\frac{\alpha}{\beta}}, -\sqrt{\frac{\beta}{\alpha}} \right),$$

with eigenvalue $\lambda_2 := 1 - \alpha - \beta$. We normalized $f_2$ so $\|f_2\|_{\pi}^2 = 1$. 
Example: two-state chain II

The spectral decomposition is therefore

\[ P^t D_\pi^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + (1 - \alpha - \beta)^t \begin{pmatrix} \frac{\alpha}{\beta} & -1 \\ -1 & \frac{\beta}{\alpha} \end{pmatrix}. \]

Put differently,

\[ P^t = \begin{pmatrix} \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \\ \frac{\alpha}{\alpha + \beta} & \frac{\beta}{\alpha + \beta} \end{pmatrix} + (1 - \alpha - \beta)^t \begin{pmatrix} \frac{\alpha}{\alpha + \beta} & -\frac{\alpha}{\alpha + \beta} \\ -\frac{\alpha}{\alpha + \beta} & \frac{\beta}{\alpha + \beta} \end{pmatrix}. \]

(Note for instance that the case \( \alpha + \beta = 1 \) corresponds to a rank-one \( P \), which immediately converges.)
Example: two-state chain III

Assume $\beta \geq \alpha$. Then

$$d(t) = \max_x \frac{1}{2} \sum_y |P^t(x, y) - \pi(y)| = \frac{\beta}{\alpha + \beta} |1 - \alpha - \beta|^t.$$ 

As a result,

$$t_{\text{mix}}(\varepsilon) = \left\lfloor \frac{\log \left( \varepsilon \frac{\alpha + \beta}{\beta} \right)}{\log |1 - \alpha - \beta|} \right\rfloor = \left\lfloor \frac{\log \varepsilon^{-1} - \log \left( \frac{\alpha + \beta}{\beta} \right)}{\log |1 - \alpha - \beta|^{-1}} \right\rfloor.$$
Recall:

**Theorem**

Let \( \{f_j\}_{j=1}^n \) be the eigenfunctions of a reversible and irreducible transition matrix \( P \) with corresponding eigenvalues \( \{\lambda_j\}_{j=1}^n \), as defined previously. Assume \( \lambda_1 \geq \cdots \geq \lambda_n \). We have the decomposition

\[
P^t(x, y) = 1 + \sum_{j=2}^n f_j(x)f_j(y)\lambda_j^t.
\]
Spectral gap

From the spectral decomposition, the speed of convergence of $P^t(x, y)$ to $\pi(y)$ is governed by the largest eigenvalue of $P$ not equal to 1.

**Definition (Spectral gap)**

The *absolute spectral gap* is $\gamma_* := 1 - \lambda_*$ where $\lambda_* := |\lambda_2| \lor |\lambda_n|$. The *spectral gap* is $\gamma := 1 - \lambda_2$.

Note that the eigenvalues of the lazy version $\frac{1}{2}P + \frac{1}{2}I$ of $P$ are $\left\{\frac{1}{2}(\lambda_j + 1)\right\}_{j=1}^n$ which are all nonnegative. So, there, $\gamma_* = \gamma$.

**Definition (Relaxation time)**

The *relaxation time* is defined as

$$t_{rel} := \gamma_*^{-1}.$$
Example continued: two-state chain

There two cases:

- $\alpha + \beta \leq 1$: In that case the spectral gap is $\gamma = \gamma^* = \alpha + \beta$ and the relaxation time is $t_{\text{rel}} = 1/(\alpha + \beta)$.

- $\alpha + \beta > 1$: In that case the spectral gap is $\gamma = \gamma^* = 2 - \alpha - \beta$ and the relaxation time is $t_{\text{rel}} = 1/(2 - \alpha - \beta)$. 
**Theorem**

Let $P$ be reversible, irreducible, and aperiodic with stationary distribution $\pi$. Let $\pi_{\min} = \min_x \pi(x)$. For all $\varepsilon > 0$, 

$$(t_{\text{rel}} - 1) \log \left(\frac{1}{2\varepsilon}\right) \leq t_{\text{mix}}(\varepsilon) \leq \log \left(\frac{1}{\varepsilon \pi_{\min}}\right) t_{\text{rel}}.$$ 

**Proof:** We start with the upper bound. By the lemma, it suffices to find $t$ such that $s(t) \leq \varepsilon$. By the spectral decomposition and Cauchy-Schwarz, 

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \lambda_*^t \sum_{j=2}^n |f_j(x)f_j(y)| \leq \lambda_*^t \sqrt{\sum_{j=2}^n f_j(x)^2 \sum_{j=2}^n f_j(y)^2}.$$ 

By our previous lemma, $\sum_{j=2}^n f_j(x)^2 \leq \pi(x)^{-1}$. Plugging this back above, 

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \lambda_*^t \sqrt{\pi(x)^{-1}\pi(y)^{-1}} \leq \frac{\lambda_*^t}{\pi_{\min}} = \frac{(1 - \gamma_*)^t}{\pi_{\min}} \leq e^{-\gamma_* t}.$$
Mixing time v. relaxation time II

The r.h.s. is less than $\varepsilon$ when $t \geq \log \left( \frac{1}{\varepsilon \pi_{\min}} \right) t_{\text{rel}}$.

For the lower bound, let $f_*$ be an eigenfunction associated with an eigenvalue achieving $\lambda_* := |\lambda_2| \lor |\lambda_n|$. Let $z$ be such that $|f_*(z)| = \|f_*\|_\infty$. By our previous lemma, $\sum_y \pi(y) f_*(y) = 0$. Hence

$$
\left| \lambda_* f_*(z) \right| = |P^t f_*(z)| = \left| \sum_y [P^t(z, y)f_*(y) - \pi(y)f_*(y)] \right| \\
\leq \|f_*\|_\infty \sum_y |P^t(z, y) - \pi(y)| \leq \|f_*\|_\infty 2d(t),
$$

so $d(t) \geq \frac{1}{2} \lambda_*^t$. When $t = t_{\text{mix}}(\varepsilon)$, $\varepsilon \geq \frac{1}{2} \lambda_*^{t_{\text{mix}}(\varepsilon)}$. Therefore

$$
t_{\text{mix}}(\varepsilon) \left( \frac{1}{\lambda_*} - 1 \right) \geq t_{\text{mix}}(\varepsilon) \log \left( \frac{1}{\lambda_*} \right) \geq \log \left( \frac{1}{2\varepsilon} \right).
$$

The result follows from $\left( \frac{1}{\lambda_*} - 1 \right)^{-1} = \left( \frac{1-\lambda_*}{\lambda_*} \right)^{-1} = \left( \frac{\gamma_*}{1-\gamma_*} \right)^{-1} = t_{\text{rel}} - 1$.  \[\blacksquare\]
1. Review

2. Bounding the mixing time via the spectral gap

3. Applications: random walk on cycle and hypercube

4. Infinite networks
Consider simple random walk on an $n$-cycle. That is, $V \coloneqq \{0, 1, \ldots, n - 1\}$ and $P(x, y) = 1/2$ if and only if $|x - y| = 1 \mod n$.

**Lemma (Eigenbasis on the cycle)**

For $j = 0, \ldots, n - 1$, the function

$$f_j(x) := \cos\left(\frac{2\pi jx}{n}\right), \quad x = 0, 1, \ldots, n - 1,$$

is an eigenfunction of $P$ with eigenvalue

$$\lambda_j := \cos\left(\frac{2\pi j}{n}\right).$$
Proof: Note that, for all $i$, $x$,

$$
\sum_y P(x, y) f_j(y) = \frac{1}{2} \left[ \cos \left( \frac{2\pi j (y - 1)}{n} \right) + \cos \left( \frac{2\pi j (y + 1)}{n} \right) \right]
$$

$$
= \frac{1}{2} \left[ \frac{e^{i \frac{2\pi j (y - 1)}{n}} + e^{-i \frac{2\pi j (y - 1)}{n}}}{2} + \frac{e^{i \frac{2\pi j (y + 1)}{n}} + e^{-i \frac{2\pi j (y + 1)}{n}}}{2} \right]
$$

$$
= \frac{1}{2} \left[ \frac{e^{i \frac{2\pi j y}{n}} + e^{-i \frac{2\pi j y}{n}}}{2} \right] \left[ \frac{e^{i \frac{2\pi j y}{n}} + e^{-i \frac{2\pi j y}{n}}}{2} \right]
$$

$$
= \cos \left( \frac{2\pi j y}{n} \right) \left[ \cos \left( \frac{2\pi j y}{n} \right) \right]
$$

$$
= \cos \left( \frac{2\pi j}{n} \right) f_j(y).
$$
Theorem (Relaxation time on the cycle)

**The relaxation time for lazy simple random walk on the cycle is**

\[ t_{\text{rel}} = \frac{2}{1 - \cos \left( \frac{2\pi}{n} \right)} = \Theta(n^2). \]

*Proof:* The eigenvalues are

\[ \frac{1}{2} \left[ \cos \left( \frac{2\pi j}{n} \right) + 1 \right]. \]

The spectral gap is therefore \( \frac{1}{2} (1 - \cos \left( \frac{2\pi}{n} \right)) \). By a Taylor expansion,

\[ 1 - \cos \left( \frac{2\pi}{n} \right) = \frac{4\pi^2}{n^2} + O(n^{-4}). \]

Since \( \pi_{\text{min}} = 1/n \), we get \( t_{\text{mix}}(\varepsilon) = O(n^2 \log n) \) and \( t_{\text{mix}}(\varepsilon) = \Omega(n^2) \). We showed before that in fact \( t_{\text{mix}}(\varepsilon) = \Theta(n^2) \).
Random walk on the cycle IV

In this case, a sharper bound can be obtained by working directly with the spectral decomposition. By Jensen’s inequality,

\[ 4\|P^t(x, \cdot) - \pi(\cdot)\|_{TV}^2 = \left\{ \sum_y \pi(y) \left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \right\}^2 \leq \sum_y \pi(y) \left( \frac{P^t(x, y)}{\pi(y)} - 1 \right)^2 \]

\[ = \left\| \sum_{j=2}^n \lambda_j^t f_j(x) f_j \right\|_{\pi}^2 = \sum_{j=2}^n \lambda_j^{2t} f_j(x)^2. \]

The last sum does not depend on \( x \) by symmetry. Summing over \( x \) and dividing by \( n \), which is the same as multiplying by \( \pi(x) \), gives

\[ 4\|P^t(x, \cdot) - \pi(\cdot)\|_{TV}^2 \leq \sum_x \pi(x) \sum_{j=2}^n \lambda_j^{2t} f_j(x)^2 = \sum_{j=2}^n \lambda_j^{2t} \sum_x \pi(x) f_j(x)^2 = \sum_{j=2}^n \lambda_j^{2t}, \]

where we used that \( \|f_j\|_{\pi}^2 = 1 \).
Consider the non-lazy chain with \( n \) odd. We get

\[
4d(t)^2 \leq \sum_{j=2}^{n} \cos \left( \frac{2\pi j}{n} \right)^{2t} = 2 \sum_{j=1}^{(n-1)/2} \cos \left( \frac{\pi j}{n} \right)^{2t}.
\]

For \( x \in [0, \pi/2) \), \( \cos x \leq e^{-x^2/2} \). (Indeed, let \( h(x) = \log(e^{x^2/2} \cos x) \). Then \( h'(x) = x - \tan x \leq 0 \) since \( (\tan x)' = 1 + \tan^2 x \geq 1 \) for all \( x \) and \( \tan 0 = 0 \). So \( h(x) \leq h(0) = 0 \).) Then

\[
4d(t)^2 \leq 2^{(n-1)/2} \sum_{j=1}^{(n-1)/2} \exp \left( -\frac{\pi^2 j^2}{n^2} t \right) \leq 2 \exp \left( -\frac{\pi^2}{n^2} t \right) \sum_{j=1}^{\infty} \exp \left( -\frac{\pi^2 (j^2 - 1)}{n^2} t \right)
\]

\[
\leq 2 \exp \left( -\frac{\pi^2}{n^2} t \right) \sum_{\ell=0}^{\infty} \exp \left( -\frac{3\pi^2 t}{n^2} \ell \right) = \frac{2 \exp \left( -\frac{\pi^2}{n^2} t \right)}{1 - \exp \left( -\frac{3\pi^2 t}{n^2} \right)},
\]

where we used that \( j^2 - 1 \geq 3(j - 1) \) for all \( j = 1, 2, 3, \ldots \). So \( t_{\text{mix}}(\varepsilon) = O(n^2) \).
Consider simple random walk on the hypercube 

\[ V := \{-1, +1\}^n \] 

where \( x \sim y \) if \( \|x - y\|_1 = 1 \). For \( J \subseteq [n] \), we let

\[ \chi_J(x) = \prod_{j \in J} x_j, \quad x \in V. \]

These are called parity functions.

Lemma (Eigenbasis on the hypercube)

For all \( J \subseteq [n] \), the function \( \chi_J \) is an eigenfunction of \( P \) with eigenvalue

\[ \lambda_J := \frac{n - 2|J|}{n}. \]
Proof: For $x \in V$ and $i \in [n]$, let $x^{[i]}$ be $x$ where coordinate $i$ is flipped. Note that, for all $J$, $x$,

\[ \sum_y P(x, y) \chi_J(y) = \sum_{i=1}^{n} \frac{1}{n} \chi_J(x^{[i]}) = \frac{n - |J|}{n} \chi_J(x) - \frac{|J|}{n} \chi_J(x) = \frac{n - 2|J|}{n} \chi_J(x). \]
**Theorem (Relaxation time on the hypercube)**

*The relaxation time for lazy simple random walk on the hypercube is*

\[ t_{\text{rel}} = n. \]

*Proof:* The eigenvalues are \( \frac{n-|J|}{n} \) for \( J \subseteq [n] \). The spectral gap is \( \gamma^* = \gamma = 1 - \frac{n-1}{n} = \frac{1}{n} \).

Because \( |V| = 2^n \), \( \pi_{\min} = 1/2^n \). Hence we have \( t_{\text{mix}}(\varepsilon) = O(n^2) \) and \( t_{\text{mix}}(\varepsilon) = \Omega(n) \). We have shown before that in fact \( t_{\text{mix}}(\varepsilon) = \Theta(n \log n) \).
As we did for the cycle, we obtain a sharper bound by working directly with the spectral decomposition. By the same argument,

$$4d(t)^2 \leq \sum_{J \neq \emptyset} \lambda_J^{2t}.$$ 

Consider the lazy chain again. Then

$$4d(t)^2 \leq \sum_{J \neq \emptyset} \left( n - \frac{|J|}{n} \right)^{2t} = \sum_{\ell=1}^{n} \binom{n}{\ell} \left( 1 - \frac{\ell}{n} \right)^{2t} \leq \sum_{\ell=1}^{n} \binom{n}{\ell} \exp \left( -\frac{2t\ell}{n} \right)$$

$$= \left( 1 + \exp \left( -\frac{2t}{n} \right) \right)^n - 1.$$ 

So $t_{\text{mix}}(\varepsilon) \leq \frac{1}{2} n \log n + O(n).$
Review

Bounding the mixing time via the spectral gap

Applications: random walk on cycle and hypercube

Infinite networks
Some remarks about infinite networks I

Remark (Positive recurrent case)

The previous results cannot in general be extended to infinite networks. Suppose $P$ is irreducible, aperiodic and positive recurrent. Then it can be shown that, if $\pi$ is the stationary distribution, then for all $x$

$$\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \to 0,$$

as $t \to +\infty$. However, one needs stronger conditions on $P$ than reversibility for the spectral theorem to apply (in a form similar to what we used above), e.g., compactness (that is, $P$ maps bounded sets to relatively compact sets, i.e. sets whose closure is compact).
Example (A positive recurrent chain whose $P$ is not compact)

For $p < 1/2$, let $(X_t)$ be the birth-death chain with $V := \{0, 1, 2, \ldots\}$, $P(0, 0) := 1 - p$, $P(0, 1) = p$, $P(x, x + 1) := p$ and $P(x, x - 1) := 1 - p$ for all $x \geq 1$, and $P(x, y) := 0$ if $|x - y| > 1$. As can be checked by direct computation, $P$ is reversible with respect to the stationary distribution $\pi(x) = (1 - \gamma)\gamma^x$ for $x \geq 0$ where $\gamma := \frac{p}{1-p}$. For $j \geq 1$, define $g_j(x) := \pi(j)^{-1/2}1_{\{x=j\}}$. Then $\|g_j\|^2_\pi = 1$ for all $j$ so $\{g_j\}_j$ is bounded in $\ell^2(V, \pi)$. On the other hand,

$$Pg_j(x) = p\pi(j)^{-1/2}1_{\{x=j-1\}} + (1 - p)\pi(j)^{-1/2}1_{\{x=j+1\}}.$$
So

\[ \| Pg_j \|_\pi^2 = p^2 \pi(j)^{-1} \pi(j - 1) + (1 - p)^2 \pi(j)^{-1} \pi(j + 1) = 2p(1 - p). \]

Hence \( \{ Pg_j \}_j \) is also bounded. However, for \( j > \ell \)

\[ \| Pg_j - Pg_\ell \|_\pi^2 \geq (1 - p)^2 \pi(j)^{-1} \pi(j + 1) + p^2 \pi(\ell)^{-1} \pi(\ell - 1) = 2p(1 - p). \]

So \( \{ Pg_j \}_j \) does not have a converging subsequence and therefore is not relatively compact.
Most random walks on infinite networks we have encountered so far were transient or null recurrent. In such cases, there is no stationary distribution to converge to. In fact:

**Theorem**

If $P$ is an irreducible chain which is either transient or null recurrent, we have for all $x, y$

$$\lim_{t \to \infty} P^t(x, y) = 0.$$ 

**Proof:**
Consider the null recurrent case. Fix $x \in V$. We observe first that:

- It suffices to show that $P^t(x, x) \to 0$. Indeed, by irreducibility, for any $y$ there is $s > 0$ such that $P^s(x, y) > 0$. So $P^{t+s}(x, x) \geq P^t(x, y)P^s(y, x)$ so $P^t(x, x) \to 0$ implies $P^t(x, y) \to 0$.

- Let $\ell = \gcd\{t : P^t(x, x) > 0\}$. As $P^t(x, x) = 0$ for any $t$ that is not a multiple of $\ell$, it suffices to consider the transition matrix $\tilde{P} := P^\ell$. That corresponds to “looking at the chain” at times $k\ell$, $k \geq 0$. We restrict the state space to $\tilde{V} := \{y \in V : \exists s \geq 0, \tilde{P}^s(x, y) > 0\}$. Let $(\tilde{X}_t)$ be the corresponding chain, and let $\tilde{P}_x$ and $\tilde{E}_x$ be the corresponding measure and expectation. Clearly we still have $\tilde{P}_x[\tau_x^+ < +\infty] = 1$ and $\tilde{E}_x[\tau_x^+] = +\infty$ because returns to $x$ under $P$ can only happen at times that are multiples of $\ell$. The reason to consider $\tilde{P}$ is that it is irreducible and aperiodic, as we show next. Note that the irreducibility of $\tilde{P}$ also implies that $\tilde{P}$ is null recurrent.
We first show that $\tilde{P}$ is irreducible. By definition of $\tilde{V}$, it suffices to prove that, for any $w \in \tilde{V}$, there exists $s \geq 0$ such that $\tilde{P}^s(w, x) > 0$. Indeed that then implies that all states in $\tilde{V}$ communicate through $x$. Let $r \geq 0$ be such that $\tilde{P}^r(x, w) > 0$. If it were the case that $\tilde{P}^s(w, x) = 0$ for all $s \geq 0$, that would imply that $\tilde{P}^x_x[\tau_x^+ = +\infty] > \tilde{P}^r(x, w) > 0$—a contradiction.

We claim further that $\tilde{P}$ is aperiodic. Indeed, if $\tilde{P}$ had period $k > 1$, then the greatest common divisor of $\{t : P^t(x, x) > 0\}$ would be $\geq k\ell$—a contradiction.

The chain $(\tilde{X}_t)$ has stationary measure

$$\mu_x(w) = \tilde{E}_x \left[ \sum_{s=0}^{\tau_x^+ - 1} \mathbb{1}_{\{\tilde{X}_s = w\}} \right] < +\infty,$$

which satisfies $\mu_x(x) = 1$ by definition and $\sum_w \mu_x(w) = +\infty$ by null recurrence.
Lemma

For any probability distribution $\nu$ on $\tilde{V}$,

$$\limsup_t \nu \tilde{P}^t(x) \leq \limsup_t \tilde{P}^t(x, x).$$

Proof: Since $\tilde{P}_\nu [\tau_x^+ = +\infty] = 0$, for any $\varepsilon > 0$ there is $N$ such that $\tilde{P}_\nu [\tau_x^+ > N] \leq \varepsilon$. So,

$$\limsup_t \nu \tilde{P}^t(x) \leq \varepsilon + \limsup_t \sum_{s=1}^N \tilde{P}_\nu [\tau_x^+ = s] \tilde{P}^{t-s}(x, x) \leq \varepsilon + \limsup_t \tilde{P}^t(x, x).$$

Since $\varepsilon$ is arbitrary, the result follows.
For $M \geq 0$, let $F \subseteq \tilde{V}$ be a finite set such that $\mu_x(F) \geq M$. Consider the conditional distribution

$$\nu_F(W) := \frac{\mu_x(W \cap F)}{\mu_x(F)}.$$ 

**Lemma**

$$(\nu_F \tilde{P}^t)(x) \leq \frac{1}{M}, \quad \forall t$$

**Proof:** Indeed

$$ (\nu_F \tilde{P}^t)(x) \leq \frac{(\mu_x \tilde{P}^t)(x)}{\mu_x(F)} = \frac{\mu_x(x)}{\mu_x(F)} \leq \frac{1}{M},$$

by stationarity.
Because $F$ is finite and $Q$ is aperiodic, there is $m$ such that $\tilde{P}^m(x, z) > 0$ for all $z \in F$. Then we can choose $\delta > 0$ such that

$$\tilde{P}^m(x, \cdot) = \delta \nu_F(\cdot) + (1 - \delta) \nu_0(\cdot),$$

for some probability measure $\nu_0$. Then

$$\limsup_t \tilde{P}^t(x, x) = \delta \limsup_t (\nu_F \tilde{P}^{t-m})(x) + (1 - \delta) \limsup_t (\nu_0 \tilde{P}^{t-m})(x) \leq \frac{\delta}{M} + (1 - \delta) \limsup_t \tilde{P}^t(x, x).$$

Rearranging gives $\limsup_t \tilde{P}^t(x, x) \leq 1/M$. Since $M$ is arbitrary, this concludes the proof.
Basic definitions I

Let \((X_t)\) be an irreducible Markov chain on a countable state space \(V\) with transition matrix \(P\) and stationary measure \(\pi > 0\). As we did in the finite case, we let \((Pf)(x) := \sum_y P(x, y)f(y)\). Let \(\ell_0(V)\) be the set of real-valued functions on \(V\) with finite support and let \(\ell^2(V, \pi)\) be the Hilbert space of real-valued functions \(f\) with \(\|f\|_\pi^2 := \sum_x \pi(x)f(x)^2 < +\infty\) equipped with the inner product

\[
\langle f, g \rangle_\pi := \sum_{x \in V} \pi(x)f(x)g(x).
\]

Then \(P\) maps \(\ell^2(V, \pi)\) to itself. In fact, we have the stronger statement:
Lemma

For any $f \in \ell^2(\mathcal{V}, \pi)$, $Pf$ is well-defined and further we have $\|Pf\|_\pi \leq \|f\|_\pi$.

Proof: Note that by Cauchy-Schwarz, Fubini and stationarity

$$\sum_x \pi(x) \left[ \sum_y P(x,y)|f(y)| \right]^2 \leq \sum_x \pi(x) \sum_y P(x,y)f(y)^2$$

$$= \sum_\pi \sum_x \pi(x)P(x,y)f(y)^2$$

$$= \sum_y \pi(y)f(y)^2 = \|f\|_\pi^2 < +\infty.$$

This shows that $Pf$ is well-defined since $\pi > 0$. Applying the same argument to $\|Pf\|_\pi^2$ gives the inequality.
We consider the operator norm

\[ \| P \|_{\pi} = \sup \left\{ \frac{\| Pf \|_{\pi}}{\| f \|_{\pi}} : f \in \ell^2(V, \pi), \ f \neq 0 \right\}, \]

and note that by the lemma \( \| P \|_{\pi} \leq 1 \). Note that, if \( V \) is finite or more generally if \( \pi \) is summable, then we have \( \| P \|_{\pi} = 1 \) since we can take \( f \equiv 1 \) above in that case.
Lemma

If in addition $P$ is reversible with respect to $\pi$, then $P$ is self-adjoint on $\ell^2(V, \pi)$, that is,

$$\langle f, Pg \rangle_\pi = \langle Pf, g \rangle_\pi \quad \forall f, g \in \ell^2(V, \pi).$$

Proof: First consider $f, g \in \ell_0(V)$. Then by reversibility

$$\langle f, Pg \rangle_\pi = \sum_{x,y} \pi(x)P(x, y)f(x)g(y) = \sum_{x,y} \pi(y)P(y, x)f(x)g(y) = \langle Pf, g \rangle_\pi.$$

Because $\ell^0(V)$ is dense in $\ell^2(V, \pi)$ (just truncate) and the bilinear form above is continuous in $f$ and $g$ (because $|\langle f, Pg \rangle_\pi| \leq ||P||_\pi ||f||_\pi ||g||_\pi$ by Cauchy-Schwarz and the definition of the operator norm) the result follows for $f, g \in \ell^2(V, \pi)$. □
Rayleigh quotient I

For a reversible $P$, we have the following characterization of the operator norm in terms of the so-called Rayleigh quotient.

**Theorem**

Let $P$ be irreducible and reversible with respect to $\pi > 0$. Then

$$\|P\|_\pi = \sup \left\{ \frac{\langle f, P f \rangle_\pi}{\langle f, f \rangle_\pi} : f \in \ell_0(V), f \neq 0 \right\}.$$

**Proof:** Let $\lambda_1$ be the r.h.s. above. By Cauchy-Schwarz $|\langle f, P f \rangle_\pi| \leq \|f\|_\pi \|P f\|_\pi$. That gives $\lambda_1 \leq \|P\|_\pi$ by dividing both sides by $\|f\|_\pi^2$. 
In the other direction, note that for a self-adjoint operator $P$ we have the following “polarization identity”

$$\langle Pf, g \rangle_\pi = \frac{1}{4} \left[ \langle P(f + g), f + g \rangle_\pi - \langle P(f - g), f - g \rangle_\pi \right],$$

which can be checked by expanding the r.h.s. Note that if $\langle f, Pf \rangle_\pi \leq \lambda_1 \langle f, f \rangle_\pi$ for all $f \in \ell_0(V)$ then the same holds for all $f \in \ell^2(V, \pi)$ because $\ell_0(V)$ is dense in $\ell^2(V, \pi)$. So for any $f, g \in \ell^2(V, \pi)$

$$|\langle Pf, g \rangle_\pi| \leq \frac{\lambda_1}{4} \left[ \langle f + g, f + g \rangle_\pi + \langle f - g, f - g \rangle_\pi \right] = \lambda_1 \frac{\langle f, f \rangle_\pi + \langle g, g \rangle_\pi}{2}.$$

Taking $g := Pf \|f\|_\pi / \|Pf\|_\pi$ gives

$$\|Pf\|_\pi \|f\|_\pi \leq \lambda_1 \|f\|_\pi^2,$$

or $\|P\|_\pi \leq \lambda_1$. \hfill \blacksquare
Definition

Let $P$ be irreducible. The *spectral radius* of $P$ is defined as

$$\rho(P) := \limsup_{t} P^t(x, y)^{1/t},$$

which does not depend on $x, y$.

To see that the lim sup does not depend on $x, y$, let $u, v, x, y \in V$ and $k, m \geq 0$ such that $P^m(u, x) > 0$ and $P^k(y, v)$. Then

$$P^{t+m+k}(u, v)^{1/(t+m+k)}$$

$$\geq (P^m(u, x)P^t(x, y)P^k(y, v))^{1/(t+m+k)}$$

$$\geq P^m(u, x)^{1/(t+m+k)}P^t(x, y)^{1/t}P^k(y, v)^{1/(t+m+k)},$$

which shows that $\limsup_{t} P^t(u, v)^{1/t} \geq \limsup_{t} P^t(x, y)^{1/t}$ for all $u, v, x, y$. 
In the positive recurrent case (for instance if the chain is finite), we have $P^t(x, y) \to \pi(y) > 0$ and so $\rho(P) = 1 = \|P\|_{\pi}$. The equality between $\rho(P)$ and $\|P\|_{\pi}$ holds in general for reversible chains.

**Theorem**

Let $P$ be irreducible and reversible with respect to $\pi > 0$. Then

$$\rho(P) = \|P\|_{\pi}.$$  

Moreover for all $t$

$$P^t(x, y) \leq \sqrt{\frac{\pi(y)}{\pi(x)}} \|P\|_{\pi}^t.$$
Proof: Because $P$ is self-adjoint and $\|\delta_z\|_\pi^2 = \pi(z) \leq 1$, by Cauchy-Schwarz
\[
\pi(x) P^t(x, y) = \langle \delta_x, P^t \delta_y \rangle_\pi \leq \|P\|_\pi^t \|\delta_x\|_\pi \|\delta_y\|_\pi = \|P\|_\pi^t \sqrt{\pi(x) \pi(y)}.
\]
Hence $P^t(x, y) \leq \sqrt{\frac{\pi(y)}{\pi(x)}} \|P\|_\pi^t$ and further $\rho(P) \leq \|P\|_\pi$.

For the other direction, by self-adjointness and Cauchy-Schwarz, for any $f \in \ell^2(V, \pi)$
\[
\|P^{t+1} f\|_\pi^2 = \langle P^{t+1} f, P^{t+1} f \rangle_\pi = \langle P^{t+2} f, P^t f \rangle_\pi \leq \|P^{t+2} f\|_\pi \|P^t f\|_\pi,
\]
or
\[
\frac{\|P^{t+1} f\|_\pi}{\|P^t f\|_\pi} \leq \frac{\|P^{t+2} f\|_\pi}{\|P^{t+1} f\|_\pi}.
\]
So $\frac{\|P^{t+1} f\|_\pi}{\|P^t f\|_\pi}$ is non-decreasing and therefore has a limit $L \leq +\infty$. Moreover 
$\frac{\|P^t f\|_\pi}{\|f\|_\pi} \leq L$ so it suffices to prove $L \leq \rho(P)$. As before it suffices to prove this for $f \in \ell_0(V)$, $f \neq 0$ by a density argument.
Observe that

\[
\left( \frac{\| P^t f \|_\pi}{\| f \|_\pi} \right)^{1/t} = \left( \frac{\| Pf \|_\pi}{\| f \|_\pi} \times \cdots \times \frac{\| P^t f \|_\pi}{\| P^{t-1} f \|_\pi} \right)^{1/t} \to L,
\]

so \( L = \lim_t \| P^t f \|_\pi^{1/t} \). By self-adjointness again

\[
\| P^t f \|_\pi^2 = \langle f, P^{2t} f \rangle_\pi = \sum_{x,y} \pi(x)f(x)f(y)P^{2t}(x,y).
\]

By definition of \( \rho := \rho(P) \), for any \( \varepsilon > 0 \), there is \( t \) large enough so that \( P^{2t}(x,y) \leq (\rho + \varepsilon)^{2t} \) for all \( x, y \) in the support of \( f \). In that case,

\[
\| P^t f \|_\pi^{1/t} \leq (\rho + \varepsilon) \left( \sum_{x,y} \pi(x)|f(x)f(y)| \right)^{1/2t}.
\]

The sum on the l.h.s. is finite because \( f \) has finite support. Since \( \varepsilon \) is arbitrary, we get \( \limsup_t \| P^t f \|_\pi^{1/t} \leq \rho \).
A counter-example

In the non-reversible case, the result generally does not hold. Consider asymmetric random walk on $\mathbb{Z}$ with probability $p \in (1/2, 1)$ of going to the right. Then both $\pi_0(x) := \left( \frac{p}{1-p} \right)^x$ and $\pi_1(x) := 1$ define stationary measures, but only $\pi_0$ is reversible. Under $\pi_1$, we have $\|P\|_{\pi_1} = 1$. Indeed, let $f_n(x) := \mathbb{1}_{\{|x| \leq n\}}$ and note that

$$(Pf_n)(x) = \mathbb{1}_{\{|x| \leq n-1\}} + p\mathbb{1}_{\{x = -n-1 \text{ or } -n\}} + (1-p)\mathbb{1}_{\{x = n \text{ or } n+1\}},$$

so $\|f_n\|_{\pi_1}^2 = 2n + 1$ and $\|Pf_n\|_{\pi_1}^2 \geq 2(n - 1) + 1$. Hence $\limsup_n \frac{\|Pf_n\|_{\pi_1}}{\|f_n\|_{\pi_1}} \geq 1$.

On the other hand, $\mathbb{E}_0[X_t] = (2p - 1)t$ and $X_t$, as a sum of $t$ independent increments in $\{-1, +1\}$, is a 2-Lipschitz function. So by the Azuma-Hoeffding inequality

$$P^t(0, 0)^{1/t} \leq \mathbb{P}_0[X_t \leq 0]^{1/t} = \mathbb{P}_0[X_t - (2p - 1)t \leq -(2p - 1)t]^{1/t} \leq e^{-\frac{2(2p-1)^2t^2}{2^2t}}.$$ 

Therefore $\rho(P) \leq e^{-(2p-1)^2/2} < 1$. 

Sébastien Roch, UW–Madison

Modern Discrete Probability – Spectral Techniques
A corollary

Corollary

Let $P$ be irreducible and reversible with respect to $\pi$. If $\|P\|_{\pi} < 1$, then $P$ is transient.

Proof: By the theorem, $P^t(x, x) \leq \|P\|_{\pi}^t$ so $\sum_t P^t(x, x) < +\infty$. Because $\sum_t P^t(x, x) = \mathbb{E}_x[\sum_t 1\{X_t=x\}]$, we have that $\sum_t 1\{X_t=x\} < +\infty$, $\mathbb{P}_x$-a.s., and $(X_t)$ is transient.

This is not an if and only if. Random walk on $\mathbb{Z}^3$ is transient, yet $P^{2t}(0, 0) = \Theta(t^{-3/2})$ so $\|P\|_{\pi} = \rho(P) = 1$. 

Sébastien Roch, UW–Madison
Modern Discrete Probability – Spectral Techniques