Convergence Rates and Efficiency Dominance of Markov Chains
Supervisor: *Jeffrey Rosenthal*

Geoffrey Salmon

September 2023

1 Introduction

This paper deals with important inequalities found in the study of convergence rates and variance bounding for Markov Chains, and Markov Chain Monte Carlo. We present results for bounding the convergence rate of Markov Chains to it's stationary distribution using total variation distance, and results on efficiency dominance of finite space Chains. We also provide a partial answer to an open-problem posed by Jeffrey Rosenthal and Radford Neal in *Efficiency of Reversible MCMC Methods: Elementary Derivations and Applications to Composite Methods*, as well prove new results on the efficiency dominance of finite chains on perpendicular subspaces of L^2. We also present bounds on the total variation distance for finite product measures. The Convergence rates of discrete-time time-homogeneous Markov Chains to their stationary distributions are explored. Sections 2-3 and 10-11 of this paper is a very thorough retelling of the results presented in *Markov Chains, Eigenvalues and Couplings* by Jeffrey Rosenthal, covering finite state space chains in sections 2-3, finite group state spaces in section 10, and general state space chains through the use of coupling in section 11. Those sections fill in nearly every gap of the paper by Jeffrey Rosenthal, making it much more accessible to audiences newer to probability and stochastic processes. Sections 4,5 and 7-9 cover the Efficiency Dominance of finite Markov Chains covered in the paper *Efficiency of Reversible MCMC Methods: Elementary Derivations and Applications to Composite Methods* by Radford Neal and Jeffrey Rosenthal, once again making results more explicit and accessible. Section 6 presents the partial answer to the open problem of finding an explicit function for which Q is more efficient than P when $Q - P$ has negative eigenvalues, and presents results converining the efficiency dominance of chains on perpendicular subspaces. Finally, section 12 provides a some final inequalities and technical lemmas, including an original bound of the variation distance of product probability measures on finite state spaces. Topics are introduced from the initial definitions, and the text is almost completely self-contained, with the only prior knowledge required being basic probability theory and linear algebra.

2 Initial Definitions

We start with the initial definitions of discrete-time, time-homogeneous Markov Chains.

Definition (Discrete-Time, Time-Homogeneous Markov Chains). *Given a measure space $(\mathcal{X}, \mathcal{M})$ with state space $\mathcal{X}$ and event space $\mathcal{M}$, a discrete-time, time-homogeneous Markov Chain is a countable (hence discrete-time) sequence of random variables $X_0, X_1, \ldots$ on $(\mathcal{X}, \mathcal{M})$ such that*

1. *there exists an initial distribution μ_0 on $(\mathcal{X}, \mathcal{M})$ such that $\mathbb{P}(X_0 \in E) = \mu_0(E)$, $\forall E \in \mathcal{M}$,*
2. *there exists a transition kernel $P(x, dy)$, such that $P(x, \cdot)$ is a probability measure on $(\mathcal{X}, \mathcal{M})$, $\forall x \in \mathcal{X}$,*
3. *where $\mathbb{P}(X_{n+1} \in E | X_n = x) = P(x, E) = \mathbb{P}(X_1 \in E | X_0 = x)$, $\forall E \in \mathcal{M}$ and $\forall x \in \mathcal{X}$, $\forall n \in \mathbb{N}$. I.e. the tranisition probabilities do not depend on the time n (hence time-homogeneous).*

Some important things to point out, is that both the initial distribution μ_0 and the transition kernel P satisfy $\int_{\mathcal{X}} \mu_0(dy) = \int_{\mathcal{X}} P(x, dy) = 1$, $\forall x \in \mathcal{X}$.

This paper is mostly concerned with finite state space Markov Chains. Luckily, in this case, our definitions can be drastically simplified.

Definition (Finite State Space Discrete-Time, Time-Homogeneous Markov Chains). *Given a non-empty finite state space $\mathcal{X}$, a Discrete-Time, Time-Homogeneous Markov Chain is a countable sequence of random variables $X_0, X_1, \ldots$ on $\mathcal{X}$ such that*

1. *there exists an initial distribution* μ_0 *on* $\mathcal{X}$ *such that* $\mu_0(x) = \mathbb{P}(X_0 = x)$, $\forall x \in \mathcal{X}$,

2. *a transition kernel* $P(x, y)$, *such that* $P(x, \cdot)$ *is a probability distribution on* $\mathcal{X}$ *for every* $x \in \mathcal{X}$,

3. *where* $\mathbb{P}(X_{n+1} = y | X_n = x) = P(x, y) = \mathbb{P}(X_1 = y | X_0 = x)$, $\forall x, y \in \mathcal{X}$, $\forall n \in \mathbb{N}$.

This is equivalent to the previous definition, taking $\mathcal{X}$ to be finite, and taking $\mathcal{M} = \mathcal{P}(\mathcal{X})$. We can make this assumption without loss of generality, as for any non-empty finite state space $\mathcal{X}$ and event space $\mathcal{M}$ on $\mathcal{X}$, we can define a new non-empty finite state space $\mathcal{X}'$ such that $\mathcal{P}(\mathcal{X}') = \mathcal{M}$. In the rest of our study of finite state space Markov Chains, we will without loss of generality always make the assumption we are working with $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$.

In this case, our earlier remark simplifies to $\sum_{y \in \mathcal{X}} \mu_0(y) = \sum_{y \in \mathcal{X}} P(x, y) = 1$. Furthermore, assuming $\mathcal{X} = \{x_0, \ldots, x_{n-1}\}$, we can express the transition kernel P as a matrix,

$$P = \begin{bmatrix} P(x_0, x_0) & \ldots & P(x_0, x_{n-1}) \\ \vdots & \ddots & \vdots \\ P(x_{n-1}, x_0) & \ldots & P(x_{n-1}, x_{n-1}) \end{bmatrix}.$$

Continuing with this linear algebraic idea, we can express the initial distribution as both a set $\{\mu_0(x_i) : x_i \in \mathcal{X}\}$ and as a vector $\mu_0 = (\mu_0(x_0), \ldots, \mu_0(x_{n-1}))$.

Given an initial distribution μ_0 and a transition kernel $P(x, dy)$ on a probability space $(\mathcal{X}, \mathcal{M})$, we can define μ_k, $\forall k \in \mathbb{N}$ to be the probability measure on $(\mathcal{X}, \mathcal{M})$ such that $\mu_k(E) = \int_{\mathcal{X}} P(x, E) \mu_{k-1}(dx)$, $\forall E \in \mathcal{M}$. This is equivalent to saying that μ_k is the probability distribution of the Markov Chain at k steps, given it's initial distribution.

Notice that if $\mathcal{X}$ is finite, this reduces to $\mu_k(E) = \sum_{x \in \mathcal{X}} P(x, E) \mu_{k-1}(x)$. Furthermore, using the notion of expressing P as a matrix and μ_k as a vector, notice that we can write this as

$$\begin{aligned}\mu_k(y) &= \sum_{x\in\mathcal{X}} P(x,y)\mu_{k-1}(x)\\ &= \sum_{x\in\mathcal{X}} p_{xy}\mu_{k-1}(x)\\ &= (\mu_{k-1}P)_y,\end{aligned}$$

so we can express $\mu_k = \mu_{k-1}P$.
Notice that as $\mu_{k-1} = \mu_{k-2}P$ too, $\mu_k = \mu_{k-1}P = \mu_{k-2}PP = \mu_{k-2}P^2$.
Repeating this k times gives us $\mu_k = \mu_0 P^k$.

Definition (Total variation distance)**.** *Let μ and v be probability distributions on the probability space $(\mathcal{X},\mathcal{M})$. Then the total variation distance between μ and v is*

$$||\mu - v||_{var} := sup_{E\in\mathcal{M}}|\mu(E) - v(E)|.$$

Sometimes the subscript will be omitted when it is understood we are talking about the total variation distance.

The total variation distance can be thought of as simply the distance between two probability distributions. So, if we are given a sequence of probability distributions, $\{\mu_k\}_{k=0}$, not necessarily from a Markov Chain, and another probability distribution v on the same space, the sequence of distributions is said to converge to v if $\lim_{k\to\infty} ||\mu_k - v||_{\text{var}} = 0$.

In our study of convergence of finite space Markov Chains, we will make alot of use of the following fact about the total variation distance on finite state spaces.

Proposition 1. *If $\mathcal{X}$ is a finite state space, and μ and v are two probability distributions on $\mathcal{X}$, then $||\mu - v|| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x) - v(x)|$.*

Proof. Notice that as $\mathcal{X}$ is finite, any $A \subseteq \mathcal{X}$ will also be finite. So,

$$||\mu - v|| := \sup_{A\subseteq\mathcal{X}}|\mu(A) - v(A)| = \max_{A\subseteq\mathcal{X}}|\sum_{x\in A}[\mu(x) - v(x)]|.$$

So, this is maximized on the set $E = \{x \in \mathcal{X} : \mu(x) \geq v(x)\}$ or $E^C = \{x \in \mathcal{X} : \mu(x) < v(x)\}$. But as μ and v are probability distributions on $\mathcal{X}$, we have

$$0 = 1 - 1 = \sum_{x \in \mathcal{X}} [\mu(x) - v(x)] = \sum_{x \in E} [\mu(x) - v(x)] + \sum_{x \in E^C} [\mu(x) - v(x)].$$

So, as $\forall x \in E$, $\mu(x) - v(x) \geq 0$ and $\forall x \in E^C$, $\mu(x) - v(x) < 0$, this is equivalent to $0 = |\sum_{x \in E}[\mu(x) - v(x)]| - |\sum_{x \in E^C}[\mu(x) - v(x)]|$. So,

$$|\sum_{x \in E} [\mu(x) - v(x)]| = |\sum_{x \in E^C} [\mu(x) - v(x)]|.$$

Thus, as $|\sum_{x \in A}[\mu(x) - v(x)]|$ is maximized when either $A = E$ or $A = E^C$, and the sum is the same using either, they both satisfy the max. Combining this with the fact that $[\mu(x) - v(x)]$ is greater or equal to zero when $x \in E$ and less than zero when $x \in E^C$, gives us

$$\begin{aligned}
2||\mu - v|| &= 2(\max_{A \subseteq \mathcal{X}} |\sum_{x \in A} \mu(x) - v(x)|) \\
&= |\sum_{x \in E} [\mu(x) - v(x)]| + |\sum_{x \in E^C} [\mu(x) - v(x)]| \\
&= \sum_{x \in E} |\mu(x) - v(x)| + \sum_{x \in E^C} |\mu(x) - v(x)| \\
&= \sum_{x \in \mathcal{X}} |\mu(x) - v(x)|.
\end{aligned}$$

□

3 Finite sample spaces and Linear Algebra

We continue on a finite state space $\mathcal{X} = \{x_0, \ldots, x_{n-1}\}$, $n \in \mathbb{N}$. Let μ_0 be an initial distribution on $\mathcal{X}$, and let P be a transition kernel on $\mathcal{X}$.

If v is a vector and $x_i \in \mathcal{X}$, then $v(x_i)$ is the i^{th} entry of v.

We start with some basic facts about the matrix P.

Proposition 2. *The stochastic matrix P has an eigenvalue $\lambda = 1$. Furthermore, the vector $v = (1, \dots, 1)$ is it's associated eigenvector.*

Proof. Let P be defined as earlier. Let $v = (1, \dots, 1)$.

$$Pv = \begin{bmatrix} P(x_0, x_0) + \cdots + P(x_0, x_{n-1}) \\ \vdots \\ P(x_{n-1}, x_0) + \cdots + P(x_{n-1}, x_{n-1}) \end{bmatrix} = v,$$

as $\forall i \in \{0, \dots, n-1\}$, $P(x_i, x_0) + \cdots + P(x_i, x_{n-1}) = 1$ (Law of Total Probability). $\square$

Notation. *Let $\lambda_0, \dots, \lambda_{n-1}$ be the eigenvalues (generalized eigenvalues in the case of a non-diagonalizable stochastic matrix) of P, and assume $\lambda_0 = 1$ is the eigenvalue found in the above proposition.*

Definition. *Let $\lambda_* := max_{1 \leq i \leq n-1} |\lambda_i|$.*

Proposition 3. *For any stochastic matrix P, $\lambda_* \leq 1$.*

Proof. Let P be a stochastic matrix. Assume λ is an eigenvalue of P. Then let v be an eigenvector of P with associated eigenvalue λ.
Let $x \in \mathcal{X}$ such that $|v(x)| \geq |v(y)|$, $\forall y \in \mathcal{X}$.
(Equivalent to choosing $x \in \mathcal{X}$ such that $|v(x)| = \max_{y \in \mathcal{X}} |v(y)|$). Then

$$\begin{aligned}
|\lambda v(x)| &= |(Pv)_x| && \text{(by assumption)} \\
&= |\sum_{y \in \mathcal{X}} P(x, y) v(y)| && \text{(by definition of } (Pv)_x\text{)} \\
&\leq \sum_{y \in \mathcal{X}} P(x, y) |v(y)| && \text{(by the triangle inequality, as } \forall x, y \in \mathcal{X},\ P(x, y) \geq 0\text{)} \\
&\leq \sum_{y \in \mathcal{X}} P(x, y) |v(x)| && \text{(by assumption)} \\
&= |v(x)| \sum_{y \in \mathcal{X}} P(x, y) \\
&= |v(x)| && \text{(by Law of Total Probability).}
\end{aligned}$$

So $|\lambda v(x)| \leq |v(x)|$, so $|\lambda| \leq 1$.
As λ was an arbitrary eigenvalue, it follows that every eigenvalue λ_i, $i \in \{0, \dots, n-1\}$, satisfies $|\lambda_i| \leq 1$.
So $\lambda_* \leq 1$. $\square$

Proposition 4. *If $P(x,y) > 0$, $\forall x, y \in \mathcal{X}$, then we get the strict inequality $\lambda_* < 1$.*

Proof. Assume that $P(x,y) > 0$, $\forall x, y \in \mathcal{X}$.
Assume initially that P is diagonalizable.
We have already seen in Proposition 2 that if $v \in \text{span}\{(1,\ldots,1)\}$, then the associated eigenvalue of v is the trivial eigenvalue.
So assume $v \notin \text{span}\{(1,\ldots,1)\}$.
If there exists $x, y \in \mathcal{X}$ such that $|v(x)| > |v(y)|$, then $\sum_{y\in\mathcal{X}} P(x,y)|v(y)| < \sum_{y\in\mathcal{X}} P(x,y)|v(x)|$, and thus the strict inequality follows, following the same line of inequalities as in the proof of Proposition 3.
If there doesn't exist $x, y \in \mathcal{X}$ such that $|v(x)| > |v(y)|$, then $\forall x, y \in \mathcal{X}$ $|v(x)| = |v(y)|$.
Then as $v \notin \text{span}\{(1,\ldots,1)\}$, there exists $x, y \in \mathcal{X}$ such that $v(x) = -v(y)$.
So $|\sum_{y\in\mathcal{X}} P(x,y)v(y)| < \sum_{y\in\mathcal{X}} P(x,y)|v(y)|$, and the strict inequality follows.

Now assume P is not diagonalizable.
If λ_0 is part of Jordan block of size greater than one, then there exists vectors not in the span of $u = (1,\ldots,1)$, such that it is an associated eigenvector of λ_0.
So, we must show the Jordan block of λ_0 cannot be greater than one, and the result will follow from the previous work on P when it was diagonalizable.
So, assume for a contradiction that the Jordan block of λ_0 is of size greater than one.
Then there exists a generalized eigenvector v such that $Pv = v + u$, where u is an ordinary eigenvector of λ_0.
As from Proposition 2 $u = (1,\ldots,1)$ is an eigenvector of λ_0, there exists v such that $Pv = v + u$ where $u = (1,\ldots,1)$.
Similarly to earlier, choose $x \in \mathcal{X}$ such that $Rev(x) \geq Rev(y)$, $\forall y \in \mathcal{X}$, where $Rev(x)$ is the real part of $v(x)$.
Again, this is equivalent to choosing $x \in \mathcal{X}$ such that $Rev(x) = \max_{y\in\mathcal{X}} Rev(y)$.

Then,

$$\begin{aligned}
1 + Rev(x) &= Re(u+v)(x) && (\text{as } u(x) = 1\ \forall x \in \mathcal{X}) \\
&= Re(Pv)_x && (\text{by definition}) \\
&= Re \sum_{y\in\mathcal{X}} P(x,y)v(y) \\
&\leq Re \sum_{y\in\mathcal{X}} P(x,y)v(x) && (\text{by assumption}) \\
&= Rev(x) \sum_{y\in\mathcal{X}} P(x,y) \\
&= Rev(x). && (\text{as } \sum_{y\in\mathcal{X}} P(x,y) = 1)
\end{aligned}$$

But obviously $1 + Rev(x) \not\leq Rev(x)$ (as $Rev(x) < \infty$, so otherwise $1 \leq 0$), so we get a contradiction, and thus the Jordan block of λ_0 must be of size 1. So, the work done earlier, showing that if $v \notin \text{span}\{(1,\ldots,1)\}$ and v was an eigenvector of P then there exists $x \in \mathcal{X}$ such that $|\lambda v(x)| < |v(x)|$ still holds, so it follows that $\lambda_* < 1$. □

Now we begin to see why λ_* is an important value for us in studying the convergence of these Chains.

Lemma 5. *If P is a stochastic matrix such that $\lambda_* < 1$ and P is diagonalizable, then there exists a unique stationary distribution, $\pi = a_0 v_0$, such that*

$$|\mu_k(x) - \pi(x)| \leq (\sum_{i=1}^{n-1} |a_i v_i(x)|)(\lambda_*)^k,$$

$\forall x \in \mathcal{X}$, where the v_i are a basis of right-eigenvectors corresponding to λ_i, and the a_i are the unique complex numbers such that $\mu_0 = a_0 v_0 + \cdots + a_{n-1} v_{n-1}$.

Proof. Assume $\lambda_* < 1$, and P is diagonalizable.
As P is diagonalizable, there exists a set of right-eigenvectors $v_0, v_1, \ldots, v_{n-1}$ corrsponding to the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ respectively, such that $\{v_0, \ldots, v_{n-1}\}$ is a basis for the vector space.
As μ_0 is a vector in this vector-space, there exists unique $a_0, \ldots, a_{n-1} \in \mathbb{C}$ such that $\mu_0 = a_0 v_0 + \cdots + a_{n-1} v_{n-1}$.

So, as $\mu_k = \mu_0 P^k$,

$$\begin{aligned}\mu_k &= \mu_0 P^k = (a_0 v_0 + \cdots + a_{n-1} v_{n-1}) P^k \\ &= a_0 v_0 P^k + \cdots + a_{n-1} v_{n-1} P^k \\ &= a_0 v_0 \lambda_0^k + \cdots + a_{n-1} v_{n-1} \lambda_{n-1}^k.\end{aligned}$$

As $\lambda_0 = 1$, $\mu_k = a_0 v_0 + a_1 v_1 \lambda_1^k + \cdots + a_{n-1} v_{n-1} \lambda_{n-1}^k$.
As $\lambda_* < 1$, $|\lambda_m| \leq \lambda_* < 1$, $\forall m \in \{1, \ldots, n-1\}$. So $(\lambda_m)^k \to 0$ as $k \to \infty$, $\forall m \in \{1, \ldots, n-1\}$.
So $\mu_k \to a_0 v_0$ as $k \to \infty$.
So let $\pi = \lim_{k\to\infty} \mu_k$.
Then $\pi = \lim_{k\to\infty} \mu_k = \lim_{k\to\infty} \mu_{k-1} P = \pi P$, so $\pi = \lim_{k\to\infty} \mu_k = a_0 v_0$ is a stationary distribution.
So $\forall x \in \mathcal{X}$, we get

$$\begin{aligned}|\mu_k(x) - \pi(x)| &= |a_0 v_0(x) + \cdots + a_{n-1} v_{n-1}(x) \lambda_{n-1}^k - a_0 v_0(x)| \\ &= |\sum_{i=1}^{n-1} a_i v_i(x) (\lambda_i)^k| \\ &\leq \sum_{i=1}^{n-1} |a_i v_i(x) (\lambda_i)^k| && \text{(by the triangle ineq.)} \\ &= \sum_{i=1}^{n-1} |a_i v_i(x)| |(\lambda_i)^k| \\ &\leq (\sum_{i=1}^{n-1} |a_i v_i(x)|)(\lambda_*)^k && \text{(as } \lambda_* = \max_{1\leq i\leq n-1} |\lambda_i|\text{).}\end{aligned}$$

□

Note that even if the matrix P is not diagonalizable, a similar upper bound is achievable, only requiring keeping track of some extra terms.

This result leads us to our bound on the convergence rates of finite space Markov Chains.

Theorem 6. *Given a finite space Markov Chain, if $\lambda_* < 1$, then the Markov Chain converges geometrically quickly to the stationary distribution $\pi = a_0 v_0$,* *where the terms a_0v_0 are the same as in Lemma 5. If additionally P is diagonalizable, then $\forall k \in \mathbb{N}$,*

$$||\mu_k - \pi|| \leq \frac{1}{2}[\sum_{x\in\mathcal{X}}(\sum_{i=1}^{n-1}|a_iv_i(x)|)](\lambda_*)^k.$$

Proof. Assume that P is diagonalizable and that $\lambda_* < 1$.
Then by Lemma 5, $\forall x \in \mathcal{X}$, $|\mu_k(x) - \pi(x)| \leq (\sum_{i=1}^{n-1}|a_iv_i(x)|)(\lambda_*)^k$ where $\pi = a_0v_0$. So,

$$\begin{aligned}||\mu_k - \pi|| &= \frac{1}{2}\sum_{x\in\mathcal{X}}|\mu_k(x) - \pi(x)| \\ &\leq \frac{1}{2}\sum_{x\in\mathcal{X}}(\sum_{i=1}^{n-1}|a_iv_i(x)|)(\lambda_*)^k \\ &= \frac{1}{2}[\sum_{x\in\mathcal{X}}(\sum_{i=1}^{n-1}|a_iv_i(x)|)](\lambda_*)^k.\end{aligned}$$

As $\forall x \in \mathcal{X}$ and $\forall i \in \{1, \ldots, n-1\}$ $|a_iv_i(x)| < \infty$, because $\mathcal{X}$ is finite,

$$\frac{1}{2}[\sum_{x\in\mathcal{X}}(\sum_{i=1}^{n-1}|a_iv_i(x)|)] < \infty.$$

So $\frac{1}{2}[\sum_{x\in\mathcal{X}}(\sum_{i=1}^{n-1}|a_iv_i(x)|)](\lambda_*)^k \to 0$, as $k \to \infty$ geometrically quickly.
So, as $0 \leq ||\mu_k - \pi|| \leq \frac{1}{2}[\sum_{x\in\mathcal{X}}(\sum_{i=1}^{n-1}|a_iv_i(x)|)](\lambda_*)^k$, $||\mu_k - \pi||$ also converges at least geometrically quickly.
A similar result follows for non-diagonalizable P. □

Now we take a step in a slightly different direction, to prove a result which will be of help to us in the next chapter.

Definition. $L^2(\pi) := \{v : \langle v, v\rangle < \infty\}$, *such that* $\langle v, w\rangle := \sum_{x\in\mathcal{X}} v(x)\overline{w(x)}\pi(x)$.

Definition (Kronecker Delta)**.** *The Kronecker Delta is $\delta_{ij} : \mathcal{X} \times \mathcal{X} \to \{0, 1\}$ such that* $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$ *More generally $\mathcal{X}$ can be just about any space.*

Lemma 7. *Under the conditions of Lemma 5, if additionially the vectors* $v_0, \ldots, v_{n-1}$ *are orthonormal in* $L^2(\pi)$*, then* $\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2\pi(x) = \sum_{i=1}^{n-1} |a_i|^2|\lambda_i|^{2k} \leq (\sum_{i=1}^{n-1} |a_i|^2)(\lambda_*)^{2k}$.

Proof. From the proof of Lemma 5, we have that $|\mu_k(x) - \pi(x)| = |\sum_{i=1}^{n-1} a_i v_i(x)(\lambda_i)^k|$, where $a_1, \ldots, a_{n-1}$ and $v_1, \ldots, v_{n-1}$ are defined as earlier, but now we further assume they are orthonormal in $L^2(\pi)$.
So, $\langle v_i, v_j\rangle = \delta_{ij}$, $\forall i, j \in \{1, \ldots, n-1\}$. So as P is diagonalizable and only has real coefficients, $v_0, \ldots, v_{n-1} \in \mathbb{R}^n$, and thus we get

$$\begin{aligned}
\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2\pi(x) &= \sum_{x\in\mathcal{X}} |\sum_{i=1}^{n-1} a_i v_i(x)(\lambda_i)^k|^2\pi(x) \\
&= \sum_{x\in\mathcal{X}} [\sum_{i=1}^{n-1} a_i v_i(x)(\lambda_i)^k]^2\pi(x) \\
&= \sum_{i\neq j} a_i a_j \lambda_i^k \lambda_j^k \sum_{x\in X} v_i(x)v_j(x)\pi(x) + \sum_{i=1}^{n-1} a_i^2\lambda_i^{2k} \sum_{x\in\mathcal{X}} v_i(x)v_i(x)\pi(x) \\
&= \sum_{i\neq j} a_i a_j \lambda_i^k \lambda_j^k \sum_{x\in X} v_i(x)\overline{v_j(x)}\pi(x) + \sum_{i=1}^{n-1} a_i^2\lambda_i^{2k} \sum_{x\in\mathcal{X}} v_i(x)\overline{v_i(x)}\pi(x) \\
&\qquad\qquad (\text{as } v_0, \ldots, v_{n-1} \in \mathbb{R}^n) \\
&= \sum_{i\neq j} a_i a_j \lambda_i^k \lambda_j^k \delta ij + \sum_{i=1}^{n-1} a_i^2\lambda_i^{2k}\delta_{ii} \qquad (\text{by assumption}) \\
&= \sum_{i=1}^{n-1} a_i^2\lambda_i^{2k} \qquad (\text{by definition of } \delta_{ij}) \\
&\leq \sum_{i=1}^{n-1} a_i^2\lambda_*^{2k} \qquad (\text{by def. of } \lambda_*) \\
&= (\sum_{i=1}^{n-1} a_i^2)\lambda_*^{2k}.
\end{aligned}$$

□

Now coming back to a more general result, we prove sufficient and necessary conditions for $\lambda_* < 1$, which is in turn a sufficient condition for con-

vergence, as proved earlier.
First the conditions:

Definition. *Periodicity*

A Markov Chain with state space $\mathcal{X}$ is said to be *periodic* if $\exists \chi_1, \ldots, \chi_n \subseteq \mathcal{X}$ such that each χ_j is a subspace, and $P(x, \chi_j) = 1$, $\forall x \in \chi_{j-1}$, $\forall j \in \{2, \ldots, n\}$, and $P(x, \chi_1) = 1$, $\forall x \in \chi_n$. In words, a Markov Chain is periodic if it jumps from one subspace to another, then from that subspace to another, and so on, eventually coming back to the original subspace, then starting the cycle over again.
Conversely, a Markov Chain is said to be *aperiodic* if $\forall x \in \mathcal{X}$, $S_x \subseteq \mathbb{N}$ such that $S_x := \{k \in \mathbb{N} : P^k(x, x) > 0\}$ has a gcd of 1. In words, a Markov Chain is aperiodic if it can go from any state back to itself in a number of steps that isn't part of a pattern.

Definition. *Decomposability*

A Markov Chain with state space $\mathcal{X}$ is said to be *decomposable* if $\exists \chi_1, \chi_2 \subseteq \mathcal{X}$ such that χ_1 and χ_2 partition $\mathcal{X}$, and $\forall (x, y) \in \chi_1 \times \chi_2$ and $\chi_2 \times \chi_1$, $P(x, y) = 0$, where P is the transition kernel. In words, a Markov Chain is decomposable if there are two distinct subsets of the state space, such that the Markov Chain will never jump from one to the other.
Conversely, a Markov Chain is said to be *indecomposable* if $\forall \chi_1, \chi_2 \subseteq \mathcal{X}$ such that χ_1 and χ_2 partition $\mathcal{X}$, $\exists (x, y) \in \chi_1 \times \chi_2$ or $\exists (x, y) \in \chi_2 \times \chi_1$ such that $P(x, y) > 0$, where P is the transition kernel. In words, given any partition of $\mathcal{X}$, there always exists a way to jump from one of the subsets to the other.

Finally, the result.

Theorem 8. *Given a Finite Markov Chain, $\lambda_* < 1$ iff the Chain is indecomposable and aperiodic.*

Proof. $\Longrightarrow$: Assume for a contradiction the Markov Chain is not indecomposable, so decomposable.
Then there exists $\chi_1, \chi_2 \subseteq \mathcal{X}$ such that χ_1 and χ_2 partition $\mathcal{X}$, and $\forall (x, y) \in \chi_1 \times \chi_2$ or $\chi_2 \times \chi_1$, $P(x, y) = 0$, where P is the transition kernel.
Then define μ_i for $i = 1, 2$, such that $\forall x \in \mathcal{X}$, $\mu_i(x) = 1$ if $x \in \chi_i$ and $\mu_i(x) = 0$ if $x \notin \chi_i$. Then as $P\mu_i = \begin{bmatrix} \sum_{x \in \mathcal{X}} P(x_0, x)\mu_i(x) \\ \vdots \\ \sum_{x \in \mathcal{X}} P(x_{n-1}, x)\mu_i(x) \end{bmatrix}$, for

$0 \leq j \leq n-1$, by definition of μ_i,

$$\sum_{x \in \mathcal{X}} P(x_j, x)\mu_i(x) = \sum_{x \in \chi_i} P(x_j, x)\mu_i(x) + \sum_{x \notin \chi_i} P(x_j, x)\mu_i(x) = \sum_{x \in \chi_i} P(x_j, x).$$

As χ_1 and χ_2 partition $\mathcal{X}$, x_j corresponds to a state in χ_1 or χ_2 exclusively. In other words, $x_j \in \chi_1 \cup \chi_2$, and $x_j \notin \chi_1 \cap \chi_2$.
So, if $x_j \in \chi_i$, then $\sum_{x \notin \chi_i} P(x_j, x) = 0$, as the Markov Chain is decomposable.
So, $1 = \sum_{x \in \mathcal{X}} P(x_j, x) = \sum_{x \in \chi_i} P(x_j, x) + \sum_{x \notin \chi_i} P(x_j, x) = \sum_{x \in \chi_i} P(x_j, x)$.
If $x_j \notin \chi_i$, then $\sum_{x \in \chi_i} P(x_j, x) = 0$ as the Markov Chain is decomposable.
So, $\forall j \in \{0, \dots, n-1\}$, $\sum_{x \in \mathcal{X}} P(x_j, x)\mu_i(x) = 1$ if $x_j \in \chi_i$ and $\sum_{x \in \mathcal{X}} P(x_j, x)\mu_i(x) = 0$ if $x_j \notin \chi_i$.
So, $P\mu_i = \mu_i$, $\forall i$. So μ_1 is an eigenvector for the eigenvalue $\lambda = 1$, and μ_2 is an eigenvector for $\lambda = 1$. So $\lambda = 1$ is an eigenvalue with multiplicity two, as μ_1 and μ_2 are linearly independent. So $\exists i \in \{1, \dots, n-1\}$ such that $\lambda_i = 1$. So $1 = \lambda_i \leq \lambda_* \leq 1$ (by Proposition 3), so $\lambda_* = 1$.

Assume for a contradiction that the Markov chain is not aperiodic, so periodic.
Then by definition of periodicity, $\exists \chi_1, \dots, \chi_n \subseteq \mathcal{X}$ such that each χ_j is a subspace, and $P(x, \chi_j) = 1$, $\forall x \in \chi_{j-1}$, $\forall j \in \{2, \dots, n\}$, and $P(x, \chi_1) = 1$, $\forall x \in \chi_n$.
Note that this implies $P(x, \chi_j^C) = 0$, $\forall x \in \chi_{j-1}$, for $j \in \{2, \dots, n\}$, and $P(x, \chi_1^C) = 0$, $\forall x \in \chi_n^C$.
Then let v be the vector such that $v(x) = \exp(\frac{2\pi i j}{n})$, $\forall x \in \chi_j$, where i is the imaginary constant, $\forall j \in \{1, \dots, n\}$.
Note that although this vector seems like an odd choice, Euler's formula and the cycle like behavior of Chains with periodicity make it a natural one.

Let $\jmath \in \{1, \dots, n\}$ and $y \in \mathcal{X}$ such that $y \in \chi_j$. Then

$$\begin{aligned}
\sum_{x \in \mathcal{X}} P(y,x)v(x) &= \sum_{x \in \chi_{j+1}} P(y,x)v(x) + \sum_{x \in \chi_{j+1}^C} P(y,x)v(x) \\
&= \sum_{x \in \chi_{j+1}} P(y,x)\exp(\frac{2\pi i(j+1)}{n}) \quad \text{(as each } P(y,x) = 0 \text{ as } y \in \chi_j \text{ and } x \in \chi_j^C) \\
&= \exp(\frac{2\pi i(j+1)}{n}) \sum_{x \in \chi_{j+1}} P(y,x) \\
&= \exp(\frac{2\pi i(j+1)}{n}) P(y, \chi_{j+1}) \\
&= \exp(\frac{2\pi i(j+1)}{n}) = \exp(\frac{2\pi i}{n})\exp(\frac{2\pi ij}{n}). \qquad \text{(as } y \in \chi_j)
\end{aligned}$$

So, as the above equality holds $\forall j \in \{1, \dots, n\}$, and consequently $\forall y \in \mathcal{X}$, $Pv = \exp(2\pi i/n)P$.
And as $|\exp(2\pi i/n)| = 1$, and $v \notin \text{span}\{(1, \dots, 1)\}$, $1 \leq \lambda_* \leq 1$ by Proposition 3 , so $\lambda_* = 1$.

$\Longleftarrow$: Assume that the Markov Chain is aperiodic and indecomposable.
Further, assume that the Markov Chain has no transient states. So, $\forall x, y \in \mathcal{X}$, $\exists r_{xy} \in \{1, \dots, n\}$ such that $\mathbb{P}(X_{r_{xy}} = y | X_0 = x) > 0$.
In other words, as $\mathcal{X}$ is finite, $P^{r_{xy}}(x, y) > 0$.
Let $x \in \mathcal{X}$. Let $S_x = \{k \in \mathbb{N} : P^k(x, x) > 0\}$.
By Lemma 61 in the appendix, as the Markov Chain is aperiodic and indecomposable, $\exists k_x > 0$ such that $\forall k \geq k_x \in \mathbb{N}$, $k \in S_x$.
So, $\forall x \in \mathcal{X}$, $\exists k_x \in \mathbb{N}$ such that $\forall k \geq k_x \in \mathbb{N}$, $k \in S_x$.
So, let $k_0 = \max_{x \in \mathcal{X}}(k_x) + n$, where $n = |\mathcal{X}|$.
Then let $x, y \in \mathcal{X}$. Then as the Markov Chain has no transient states, $\exists r_{xy} \in \{1, \dots, n\}$ such that $P^{r_{xy}}(x, y) > 0$. Then,

$$
\begin{aligned}
P^{k_0}(x,y) &= \mathbb{P}(X_{k_0} = y | X_0 = x) \\
&= \sum_{j \in \mathcal{X}} \mathbb{P}(X_{k_0} = y | X_{k_0 - r_{xy}} = j) \mathbb{P}(X_{k_0 - r_{xy}} = j | X_0 = x) \\
&= \sum_{j \in \mathcal{X}} P^{k_0 - k_0 + r_{xy}}(j,y) P^{k_0 - r_{xy}}(x,j) \\
&= \sum_{j \in \mathcal{X}} P^{r_{xy}}(j,y) P^{k_0 - r_{xy}}(x,j) \\
&\geq P^{r_{xy}}(x,y) P^{k_0 - r_{xy}}(x,x) \\
&> 0,
\end{aligned}
$$

as $k_0 - r_{xy} \geq k_x$ so $P^{k_0 - r_{xy}}(x,x) > 0$, and $P^{r_{xy}}(x,y) > 0$.
In other words, the probability that we go from x to y in k_0 steps is greater than or equal to the probability that we go from x to x in $k_0 - r_{xy}$ steps then from x to y in r_{xy} steps. And as we know both the latter probabilities are greater than zero, so is the former.
As $x, y \in \mathcal{X}$ were arbitrary, $\forall x, y \in \mathcal{X}$, $P^{k_0}(x,y) > 0$. So by Proposition 4 , $\lambda_* < 1$ for P^{k_0}.
Notice that if $\lambda_* = 1$, then for the associated eigenvector, lets call μ,

$$P^{k_0}\mu = P^{k_0-1}(P\mu) = P^{k_0-1}(\pm\mu) = \cdots = \pm\mu.$$

But then obviously $\lambda_* = 1$ for P^{k_0}, so $\lambda_* < 1$ for P as well.
Now we must show that Markov Chains with transient states simplify to this solution as well.
Let $x \in \mathcal{X}$ such that x is a transient state.
Then by definition $\exists y \in \mathcal{X}$ and $r \in \mathbb{N}$ such that $P^r(x,y) > 0$ and $\forall m \in \mathbb{N} \cup \{0\}$, $P^m(y,x) = 0$.
Let $T = \{j \in \mathcal{X} : \exists m \in \mathbb{N} \cup \{0\} s.t. P^m(j,x) > 0\}$. Then obviously $y \notin T$. So,

$$
\begin{aligned}
\sum_{j \in T} |(vP^r)_j| &= \sum_{j \in T} |\sum_{l \in \mathcal{X}} v(l) P^r(l,j)| \\
&\leq \sum_{j \in T} \sum_{l \in \mathcal{X}} |v(l)| P^r(l,j) \qquad \text{(by the triangle ineq.)} \\
&= \sum_{l \in \mathcal{X}} |v(l)| \sum_{j \in T} P^r(l,j).
\end{aligned}
$$

Now if $\exists l \in T^c$ and $j \in T$ such that $P^r(l,j) > 0$, then as $j \in T$, $\exists m \in \mathbb{N} \cup \{0\}$ such that $P^m(j,x) > 0$. So, $P^{r+m}(l,x) = \sum_{k \in \mathcal{X}} P^r(l,k)P^m(k,x) \geq P^r(l,j)P^m(j,x) > 0$.
So $\exists n \in \mathbb{N} \cup \{0\}$ such that $P^n(l,x) > 0$, so $l \in T$ by definition of T.
But we have that $l \in T^c$, so we have a contradiction.
So, $\forall l \in T^c$ and $\forall j \in T$, $P^r(l,j) = 0$. So,

$$\begin{aligned}
\sum_{l \in \mathcal{X}} |v(l)| \sum_{j \in T} P^r(l,j) &= \sum_{l \in T} |v(l)| \sum_{j \in T} P^r(l,j) + \sum_{l \in T^c} |v(l)| \sum_{j \in T} P^r(l,j) \\
&= \sum_{l \in T} |v(l)| \sum_{j \in T} P^r(l,j) \\
&\text{(by the fact just proved)} \\
&= \sum_{l \in T - \{x\}} |v(l)| \sum_{j \in T} P^r(l,j) + |v(x)| \sum_{j \in T} P^r(x,j) \\
&\text{(as } x \in T, \text{ as } P^0(x,x) = 1. \text{ I.e. } x \text{ will be at } x \text{ in 0 steps.)} \\
&\leq \sum_{l \in T - \{x\}} |v(l)| + |v(x)| \sum_{j \in T} P^r(x,j) \\
&\text{(as } \sum_{j \in T} P^r(l,j) = P^r(l,T) \leq 1.) \\
&\leq \sum_{l \in T - \{x\}} |v(l)| + |v(x)|(1 - P^r(x,y)) \\
&\text{(as } y \notin T, \text{ so } 1 = P^r(x,\mathcal{X}) = \sum_{j \in T} P^r(x,j) \\
&\qquad + \sum_{j \in T^C} P^r(x,j) \geq \sum_{j \in T} P^r(x,j) + P^r(x,y).) \\
&= \sum_{l \in T} |v(l)| - P^r(x,y)|v(x)|,
\end{aligned}$$

As $P^r(x,y) > 0$, if v is an eigenvector of P with $|\lambda| = 1$ as it's associated eigenvalue, then it must satisfy $v(x) = 0$.
Then λ is also an eigenvalue of the same chain on the state space $\mathcal{X} - \{x\}$.
Thus, as x was an arbitrary transient state, every eigenvalue of P with $|\lambda| = 1$ is also an eigenvalue of the same chain with the state space $\mathcal{X} - A$, where $A = \{x \in \mathcal{X} : x \text{ is a transient state}\}$. So, it reduces to the earlier case, and the result follows. $\square$

Combining Theorems 6 and 8, we see that if a Markov Chain with a finite state space is indecomposable and apriodic, then it converges geometrically quickly to a unique stationary distribution, and we get a bound on the total variation distance for such Chains.

4 Efficiency Results

Once we know approximately how long to run a Markov Chain for it to converge, we then take samples from that time T and onwards, say $X_T, X_{T+1}, \dots,$ which allows us to approximate $\mathbb{E}_\pi(f)$ for any $f : \mathcal{X} \to \mathbb{R}$ using an important estimator,

$$\hat{f}_N = (\frac{1}{N}) \sum_{i=T}^{N+T} f(X_i). \tag{1}$$

But in order for this estimate to be accurate, not only do we need to know that each X_i is approximately distributed by π, i.e. that the total variation distance of the Markov Chain and the stationary distribution are sufficiently low, $||\mu_T - \pi|| \leq$ chosen small number, but also that the variance of $\hat{f}_N$ is sufficiently low. So when choosing between Markov Chains with the same stationary distributions, say two Markov Chains with transition kernels P and Q respectively, if $\mathbb{V}\text{ar}(\hat{f}_N) \leq \mathbb{V}\text{ar}(\hat{g}_N)$, $\forall N$, where $\hat{f}_N$ is sampled from $X_i \sim P^i$ and $\hat{g}_N$ is sampled from $Y_i \sim Q^i$, then it's obvious the Markov Chain with transition kernel P would be a better estimate most of the time. This section deals with how to know when one Markov Chain with transition kernel P is a better choice than that of Q.

We continue to assume that $\mathcal{X}$ is finite, and $\mathcal{X} = \{x_0, \dots, x_{n-1}\}$, so $|\mathcal{X}| = n$. We further assume that P and Q are the transition matrices of two Markov Chains of $\mathcal{X}$, with stationary distribution π. For the purposes of this variance comparison, we will assume the first esimate, T is equal to 1, and we take N estimates.
Now we introduce some new definitions.

Definition (Asymptotic Variance). *The asymptotic variance of the estimator* (1) *of a function $f : \mathcal{X} \to \mathbb{R}$ using the Markov Chain P, is*

$$v(f, P) := \lim_{N \to \infty} N \mathbb{V}ar(\hat{f}_N).$$

Notice that by properties of $\mathbb{V}$ar and by defintion of the estimator (1),

$$\begin{aligned}
v(f, P) &= \lim_{N\to\infty} N\mathbb{V}\text{ar}(\hat{f}_N) \\
&= \lim_{N\to\infty} N\mathbb{V}\text{ar}((1/N)\sum_{i=1}^{N} f(X_i)) \\
&= \lim_{N\to\infty} N(1/N)^2\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)) \\
&= \lim_{N\to\infty} (1/N)\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)).
\end{aligned}$$

The asymptotic method will be our tool for deciding on the better Markov Chain, which will be decided according to the following definition.

Definition (Efficiency Dominance)**.** *We say that the Markov Chain with transition matrix P efficiency dominates that with transition matrix Q if*

$$v(f, P) \leq v(f, Q) \qquad \forall f \in L_0^2(\pi),$$

where $L_0^2(\pi) \subseteq L^2(\pi)$ such that $\forall f \in L_0^2(\pi), \mathbb{E}_\pi(f) = 0$.

Recall the definition of $L^2(\pi) := \{f : \langle f, f\rangle < \infty\}$, and $\langle f, g\rangle := \sum_{i=0}^{n-1} f(x_i)\overline{g(x_i)}\pi(x_i)$. For the rest of our current discussion, we will restrict $L^2(\pi)$ to the set of functions in $\mathbb{R}^{\mathcal{X}}$. Notice that this means

$$\langle f, g\rangle := \sum_{i=0}^{n-1} f(x_i)\overline{g(x_i)}\pi(x_i) = \sum_{i=0}^{n-1} f(x_i)g(x_i)\pi(x_i),$$

as $g(x_i) \in \mathbb{R}$, $\forall x_i \in \mathcal{X}$.
We will also, for most of our further discussion, restrict to functions $f \in L^2(\pi)$ such that $\mathbb{E}_\pi(f) = 0$, denoted as $L_0^2(\pi)$, so in otherwords $L_0^2(\pi) := \{f \in \mathbb{R}^{\mathcal{X}} : \langle f, f\rangle < \infty \text{ and } \mathbb{E}_\pi(f) = 0\}$. This will not affect our results about the asymptotic variance of functions, as the asymptotic variance is expressed as a scalar multiple of variance, which is not affected by any finite mean.

Notation. *For the rest of this section, we shall refer to the transition kernel/matrix, say P of a Markov Chain as the Markov Chain. Note that this is just short for the Markov Chain whose transition matrix is P.*

Definition. *Let D be the diagonal matrix with $D(i,i) = \pi(x_i)$, $\forall i \in \{0, \dots, n-1\}$.*

Definition. *The Markov Chain P is called reversible with respect to π if $\pi(x)P(x,y) = \pi(y)P(y,x)$, $\forall x, y \in \mathcal{X}$.*

We now get into some preliminary results about the transition matrices, that will serve as the backbone for the rest of the theory.

Definition (Irreducibility)**.** *A Markov Chain with transition matrix P is irreducible if $\forall x, y \in \mathcal{X}$, $\exists r \in \mathbb{N}$ such that $P^r(x,y) > 0$.*

Proposition 9. *If P is an irreducible Markov Chain and π is it's stationary distribution, then $\pi(x) > 0$, $\forall x \in \mathcal{X}$.*

Proof. Assume for a contradiction that $x_i \in \mathcal{X}$ such that $\pi(x_i) = 0$.
As P is irreducible, $\exists x_j \in \mathcal{X}$ such that $x_j \neq x_i$ and $P(x_j, x_i) > 0$. So, as π is a stationary distribution,

$$0 = \pi(x_i) = \sum_{l=1}^{n} P(x_l, x_i)\pi(x_l) = \sum_{l \neq j} P(x_l, x_i)\pi(x_l) + P(x_j, x_i)\pi(x_j).$$

So, $\pi(x_j) = 0$. If $n = 2$, then as π is a probability distribution, $1 = \sum_{l=1}^{2} \pi(x_l) = \pi(x_j) + \pi(x_i) = 0$.
So we get a contradiction. For any $n > 2$, an inductive argument, will get that for any n, if $\pi(x) = 0$ for any $x \in \mathcal{X}$, then we must have $\pi(y) = 0$, $\forall y \in \mathcal{X}$, which contradicts that π is a probability disrtibution. $\square$

Proposition 10. *If P is reversible wrt π, then $\langle f, Pg \rangle = \langle Pf, g \rangle$, $\forall f, g \in \mathbb{R}^{\mathcal{X}}$. I.e. P is hermitian wrt $\langle \cdot, \cdot \rangle$.*

Proof. Let $f, g \in \mathbb{R}^{\mathcal{X}}$. Then

$$\begin{aligned}
\langle f, Pg \rangle &= \sum_{i=0}^{n-1} f(x_i)(Pg)(x_i)\pi(x_i) \\
&= \sum_{i=0}^{n-1} f(x_i) \sum_{k=0}^{n-1} P(x_i, x_k) g(x_k) \pi(x_i) \\
&= \sum_{i=0}^{n-1} f(x_i) \sum_{k=0}^{n-1} P(x_k, x_i) g(x_k) \pi(x_k) \\
&= \sum_{k=0}^{n-1} g(x_k) \sum_{i=0}^{n-1} P(x_i, x_k) f(x_i) \pi(x_k) \\
&= \langle Pf, g \rangle.
\end{aligned}$$

□

This next Lemma will be foundational for the rest of the theory, as it shows that if Markov Chains that are reversible wrt their stationary matrices are very simple and easy to work with.

Lemma 11. *Assume that P is irreducible and reversible wrt to π. Then P is diagonalisable with real eigenvalues, and with a set of orthonormal wrt $\langle \cdot, \cdot \rangle$ real eigenvectors.*

Proof. Notice that $\forall f, g \in \mathbb{R}^{\mathcal{X}}$, as P is reversible,

$$\begin{aligned}
\langle f, Pg \rangle :&= \sum_{i=0}^{n-1} f(x_i) \sum_{k=0}^{n-1} P(x_i, x_k) g(x_k) \pi(x_i) \\
&= \sum_{i=0}^{n-1} f(x_i) \sum_{k=0}^{n-1} P(x_k, x_i) \pi(x_k) g(x_k) \\
&=: f \cdot (DP)g.
\end{aligned}$$

So, by Proposition 10, it follows that DP is self-adjoint wrt the dot product, and thus DP is symmetric.
So, as DP is symmetric, $DP = (DP)^T$. And as D is diagonal, $(DP)^T = P^T D$, and thus $DP = P^T D$.

As P is hermitian, it is diagonalisable. So let $v \in \mathbb{R}^{\mathcal{X}}$ such that $Pv = \lambda v$, where $\lambda \in \mathbb{C}$, and v is non-zero. Then $\overline{\lambda}\overline{v}^T = \overline{v}^T P^T$. So,

$$\overline{\lambda}(\overline{v}^T Dv) = \overline{v}^T P^T Dv = \overline{v}^T DPv = \lambda(\overline{v}^T Dv).$$

So, as v is non-zero and D is diagonal with all diagonal entries non-zero by Proposition 9 as P is irreducible, $\overline{v}^T Dv$ is also non-zero. Thus $\overline{\lambda} = \lambda$, so $\lambda \in \mathbb{R}$.
As $\lambda \in \mathbb{R}$, if v is an eigenvector, then notice

$$P(Re(v)) + iP(Im(v)) = P(Re(v) + iIm(v)) = Pv = \lambda v = \lambda Re(v) + i\lambda Im(v).$$

Also, as v is an eigenvector by definition it is non-zero, so at least one of $Re(v)$ or $Im(v)$ is non-zero, and hence by the above also an eigenvector of P with associated eigenvalue λ. So, we can take all eigenvectors to be real.
Let v_i and v_j be real eigenvectors of P with eigenvectors λ_i and λ_j respectively. Then as $\lambda_i, \lambda_j \in \mathbb{R}$, and v_i and v_j are real,

$$\lambda_j(v_j^T Dv_i) = v_j^T P^T Dv_i = v_j^T DPv_i = \lambda_i v_j^T Dv_i.$$

So, if $\lambda_j \neq \lambda_i$, then $0 = v_j^T Dv_i = \langle v_j, v_i \rangle$.
Notice that if $\lambda_j = \lambda_i$, and $v_i \neq v_j$, then we have an eigenvalue λ with multiplicity greater than one. So, assuming the dimension of the eigenspace of λ is k, we can take any k orthogonal vectors in the eigenspace, and they will by the above still all be orthogonal to the other eigenvectors.
We can then just scale the orthogonal set of eigenvectors to 1 to make it orthonormal. □

Notation. *We will from now on additionally write the eigenvalues of P is descending order. I.e. $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1}$. Let $v_0, \ldots, v_{n-1} \in \mathbb{R}^{\mathcal{X}}$ be the real eigenvectors associated to the eigenvalues $\lambda_0, \ldots, \lambda_{n-1}$.*

Note that this doesn't contradict any of the earlier work, as $\lambda_0 = 1$ found in Proposition 2 is still the biggest eigenvalue, as by Proposition 3, $\lambda_* \leq \lambda_0$.

As $\{v_0, \ldots, v_{n-1}\}$ is an orthormal basis, $\forall f, g \in \mathbb{R}^{\mathcal{X}}$, we have $f = \sum_{i=0}^{n-1} a_i v_i$

where $a_i \in \mathbb{R}$, and similarly $g = \sum_{i=0}^{n-1} b_i v_i$ for $b_i \in \mathbb{R}$. So, this gives us that

$$
\begin{aligned}
\langle f, g\rangle &= \langle \sum_{i=0}^{n-1} a_i v_i, \sum_{j=0}^{n-1} b_j v_j\rangle \\
&= \sum_{i=0}^{n-1} \langle a_i v_i, \sum_{j=0}^{n-1} b_j v_j\rangle \\
&= \sum_{i=0}^{n-1} a_i \langle v_i, \sum_{j=0}^{n-1} b_j v_j\rangle \\
&= \sum_{i=0}^{n-1} a_i \sum_{j=0}^{n-1} \langle v_i, b_j v_j\rangle \\
&= \sum_{i=0}^{n-1} a_i \sum_{j=0}^{n-1} \langle v_i, b_j v_j\rangle \\
&= \sum_{i=0}^{n-1} a_i \sum_{j=0}^{n-1} b_j \langle v_i, v_j\rangle \\
&= \sum_{i=0}^{n-1} a_i \sum_{j=0}^{n-1} b_j \delta_{ij} \\
&= \sum_{i=0}^{n-1} (a_i)(b_i).
\end{aligned}
$$

Importantly, as $Pf = P(\sum_{i=0}^{n-1} a_i v_i) = \sum_{i=0}^{n-1} a_i P v_i = \sum_{i=0}^{n-1} a_i \lambda_i v_i$,

$$
\langle f, Pf\rangle = \sum_{i=0}^{n-1} (a_i)^2 \lambda_i.
$$

5 Efficiency Dominance, the Inner Product, and more Eigenvalues

Lemma 12. *If $h : \mathcal{X} \to \mathbb{R}$ is a function, then $\forall i \in \mathbb{N}$, $\mathbb{E}_{\pi,P}(h(X_i)) = \mathbb{E}_\pi(h)$.*

Proof. This follows because π is a stationary distribution. This Lemma may be very unnecessary.

When $i = 1$, it is trivial.
When $i = 2$, notice

$$\begin{aligned}\mathbb{E}_{\pi,P}(h(X_2)) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} h(y)P(x,y)\pi(x)\\ &= \sum_{y\in\mathcal{X}}\sum_{x\in\mathcal{X}} h(y)P(x,y)\pi(x)\\ &= \sum_{y\in\mathcal{X}} h(y)\sum_{x\in\mathcal{X}} P(x,y)\pi(x)\\ &= \sum_{y\in\mathcal{X}} h(y)\pi(y)\\ &= \mathbb{E}_\pi(h).\end{aligned}$$

So, we can continue inductively. Assume $\mathbb{E}_{\pi,P}(h(X_i)) = \mathbb{E}_\pi(h)$, $\forall h \in \mathbb{R}^{\mathcal{X}}$. Notice this implies
$\sum_{y\in\mathcal{X}} h(y)\sum x \in XP^i(x,y)\pi(x) = \mathbb{E}_{\pi,P}(h(X_i)) = \mathbb{E}_\pi(h) = \sum_{y\in\mathcal{X}} h(y)\pi(y)$.
Then

$$\begin{aligned}\mathbb{E}_{\pi,P}(h(X_{i+1})) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} h(y)P^{i+1}(x,y)\pi(x)\\ &= \sum_{y\in\mathcal{X}} h(y)\sum_{x\in\mathcal{X}} P^{i+1}(x,y)\pi(x)\\ &= \sum_{y\in\mathcal{X}} h(y)\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} P(z,y)P^i(x,z)\pi(x)\\ &= \sum_{y\in\mathcal{X}} h(y)\sum_{z\in\mathcal{X}} P(z,y)\sum_{x\in\mathcal{X}} P^i(x,z)\pi(x)\\ &= \sum_{y\in\mathcal{X}} h(y)\sum_{z\in\mathcal{X}} P(z,y)\pi(z)\\ &= \sum_{y\in\mathcal{X}} h(y)\pi(y)\\ &= \mathbb{E}_\pi(h).\end{aligned}$$

□

Lemma 13. $\forall f \in L_0^2(\pi)$, $(1/N)\mathbb{V}ar(\sum_{i=1}^N f(X_i)) = \langle f, f\rangle + 2\sum_{k=1}^N \frac{N-k}{N}\langle f, P^k f\rangle$.

Proof. Let $f \in L_0^2(\pi)$. Then by a theorem about the variance of a random variable, we know

$$\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)) = \mathbb{E}_{\pi,P}[(\sum_{i=1}^{N} f(X_i))^2] - [\mathbb{E}_{\pi,P}(\sum_{i=1}^{N} f(X_i))]^2.$$

But notice by properties of the expected value function, Lemma 12, and as $\mathbb{E}_\pi(f) = 0$ as $f \in L_0^2(\pi)$,

$$\begin{aligned}
\mathbb{E}_{\pi,P}(\sum_{i=1}^{N} f(X_i)) &= \sum_{i=1}^{N}[\mathbb{E}_{\pi,P}(f(X_i))] \\
&= \sum_{i=1}^{N}[\mathbb{E}_\pi(f)] \\
&= 0.
\end{aligned}$$

So, we get $\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)) = \mathbb{E}_{\pi,P}[(\sum_{i=1}^{N} f(X_i))^2]$.
Now, by expanding the square, Lemma 12, and using the linearity of $\mathbb{E}$,

$$\begin{aligned}
\mathbb{E}_{\pi,P}[(\sum_{i=1}^{N} f(X_i))^2] &= \mathbb{E}_{\pi,P}[(\sum_{i=1}^{N} f(X_i)^2) + 2\sum_{i=1}^{N}\sum_{j=1}^{i-1} f(X_i)f(X_j)] \\
&= [\sum_{i=1}^{N} \mathbb{E}_{\pi,P}(f(X_i)^2)] + 2\sum_{i=1}^{N}\sum_{j=1}^{i-1} \mathbb{E}_{\pi,P}(f(X_i)f(X_j)) \\
&= N\mathbb{E}_\pi(f^2) + 2\sum_{i=1}^{N}\sum_{j=1}^{i-1} \mathbb{E}_{\pi,P}(f(X_i)f(X_j)).
\end{aligned}$$

So, in $2\sum_{i=1}^{N}\sum_{j=1}^{i-1} \mathbb{E}_{\pi,P}(f(X_i)f(X_j))$, $i > j$ for all i and j, using $i = j + k$, we can rewrite this as

$$2\sum_{k=1}^{N}\sum_{j=1}^{N-k} \mathbb{E}_{\pi,P}(f(X_{j+k})f(X_j)).$$

As π is stationary and P is time-homogeneous, this is the same as $2\sum_{k=1}^{N}(N-k)\mathbb{E}_{\pi,P}(f(X_{j+k})f(X_j))$ for some $j \in \mathbb{N}$.
So, as $\langle f, f\rangle = \mathbb{E}_\pi(f^2)$,

$\langle f, P^k f\rangle = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} f(x)P^k(x,y)f(y)\pi(x) = \mathbb{E}_{\pi,P}(f(X_{j+k})f(X_j))$, and the above work,

$$\begin{aligned}\frac{1}{N}\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)) &= \frac{1}{N}(N\mathbb{E}_\pi(f^2) + 2\sum_{k=1}^{N}(N-k)\mathbb{E}_{\pi,P}(f(X_{j+k})f(X_j)) \\ &= \langle f, f\rangle + 2\sum_{k=1}^{N}\frac{N-k}{N}\langle f, P^k f\rangle.\end{aligned}$$

□

Theorem 14. *If P is irreducible, then P is indecomposable.*

Proof. Let P be irreducible, and assume for a contradiction that P is decomposable.
By definition of decomposability, $\exists \mathcal{X}_1, \mathcal{X}_2 \subseteq \mathcal{X}$ such that $\mathcal{X}_1$ and $\mathcal{X}_2$ partition $\mathcal{X}$, and $\forall (x,y) \in \mathcal{X}_1 \times \mathcal{X}_2$, $P(x,y) = 0$.
Let $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$ (as $\mathcal{X}_1$ and $\mathcal{X}_2$ are non-empty).
As P is irreducible, by definition, $\exists r \in \mathbb{N}$ such that $P^r(x_1, x_2) > 0$. So,

$$\begin{aligned}P^r(x_1,x_2) &= \sum_{k_1} P(x_1,k_1)P^{r-1}(k_1,x_2) \\ &= \sum_{k_1}\cdots\sum_{k_{r-1}} P(x_1,k_1)P(k_1,k_2)\cdots P(k_{r-1},x_2). \qquad (1)\end{aligned}$$

As P is decomposable and $x_1 \in \mathcal{X}_1$, if $k_1 \notin \mathcal{X}_1$ (i.e. $k_1 \in \mathcal{X}_2$) then $P(x_1, k_1) = 0$. So (1) is equal to $\sum_{k_1\in\mathcal{X}_1}\cdots\sum_{k_{r-1}} P(x_1,k_1)P(k_1,k_2)\cdots P(k_{r-1},x_2)$ (2).
As $k_1 \in \mathcal{X}_1$, if $k_2 \notin \mathcal{X}_1$ then $P(k_1, k_2) = 0$, so (2) equals
$\sum_{k_1\in\mathcal{X}_1}\sum_{k_2\in\mathcal{X}_1}\cdots\sum_{k_{r-1}} P(x_1,k_1)P(k_1,k_2)\cdots P(k_{r-1},x_2)$.
This argument continues inductively to give us

$$\sum_{k_1\in\mathcal{X}_1}\cdots\sum_{k_{r-1}\in\mathcal{X}_1} P(x_1,k_1)P(k_1,k_2)\cdots P(k_{r-1},x_2) = P^r(x_1,x_2) > 0.$$

But as P is decomposable, $\forall k_{r-1} \in \mathcal{X}_1$, $P(k_{r-1}, x_2) = 0$ as $x_2 \in \mathcal{X}_2$. But then

$$0 = \sum_{k_1\in\mathcal{X}_1}\cdots\sum_{k_{r-1}\in\mathcal{X}_1} P(x_1,k_1)P(k_1,k_2)\cdots P(k_{r-1},x_2) > 0.$$

A contradiction. □

We know show the relation between the eigenvalues and eigenvectors of P with the asymptotic variance.

Proposition 15. *If P is irreducible and reversible wrt π, then $\forall f \in L^2_0(\pi)$,*

$$v(f,P) = \sum_{i=1}^{n-1}(a_i)^2 + 2\sum_{i=1}^{n-1}(a_i)^2\frac{\lambda_i}{1-\lambda_i} = \sum_{i=1}^{n-1}(a_i)^2\frac{1+\lambda_i}{1-\lambda_i},$$

where $a_i \in \mathbb{R}$ such that $f = \sum_{i=1}^{n-1} a_i v_i$, $\forall i$, where $\{v_0, \ldots, v_{n-1}\}$ is the earlier discussed real orthonormal basis of P.

Proof. Let $f \in L^2_0(\pi)$. Then by Lemma 13, we know that

$$\frac{1}{N}\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)) = \langle f, f\rangle + 2\sum_{k=1}^{N}\frac{N-k}{N}\langle f, P^k f\rangle.$$

As P is a stochastic matrix that is irreducible and reversible wrt π, by Lemma 11, there exists an set of real orthormal eigenvectors $\{v_0, \ldots, v_{n-1}\}$ corresponding to the eigenvalues $\{\lambda_0, \ldots, \lambda_{n-1}\}$, as defined in Notation 5.
So, as $f \in L^2_0(\pi)$, $f \in \mathbb{R}^{\mathcal{X}}$, $\exists a_0, \ldots, a_{n-1} \in \mathbb{R}$ such that $f = \sum_{i=0}^{n-1} a_i v_i$. So, we get that

$$\frac{1}{N}\mathbb{V}\text{ar}(\sum_{i=1}^{N} f(X_i)) = \sum_{i=0}^{n-1}(a_i^2) + 2\sum_{k=1}^{N}\frac{N-k}{N}\sum_{i=0}^{n-1}a_i^2\lambda_i^k. \qquad (1)$$

Notice that as $f \in L^2_0(\pi)$, by definition $0 = \mathbb{E}_\pi(f) = \langle f, 1\rangle = \langle f, v_0\rangle$, so $a_0 = 0$. Furthermore, if we let $\mathbb{I}_{k\le N-1} : \mathbb{N} \to \{0,1\}$ be the indicator function for the set $\{1, \ldots, N-1\}$, then this is equivalently

$$\sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{\infty}\sum_{i=1}^{n-1}\mathbb{I}_{k\le N-1}(k)\frac{N-k}{N}a_i^2\lambda_i^k.$$

If P is also aperiodic, then as P is irreducible by Theorem 14 it is also indecomposable. So, as P is indecomposable and aperiodic, by Theorem 8

$\lambda_* < 1$. So, using the triangle inequality,

$$\begin{aligned}
\sum_{k=1}^{\infty} |\sum_{i=1}^{n-1} \mathbb{I}_{k\leq N-1}(k)\frac{N-k}{N}a_i^2\lambda_i^k| &\leq \sum_{k=1}^{\infty}\sum_{i=1}^{n-1} |\mathbb{I}_{k\leq N-1}(k)\frac{N-k}{N}a_i^2\lambda_i^k| \\
&\leq \sum_{k=1}^{\infty}\sum_{i=1}^{n-1} |a_i^2\lambda_i^k| \\
&\leq \sum_{k=1}^{\infty}\sum_{i=1}^{n-1} |a_i^2\lambda_*^k| \\
&= \sum_{k=1}^{\infty}\sum_{i=1}^{n-1} a_i^2\lambda_*^k \\
&= \sum_{k=1}^{\infty}[\lambda_*^k(\sum_{i=1}^{n-1} a_i^2)] \\
&= \sum_{i=1}^{n-1} a_i^2\frac{\lambda_*}{1-\lambda_*} \\
&< \infty,
\end{aligned}$$

as $\mathbb{I}_{k\leq N-1}\frac{N-k}{N} < 1$. So, as it is absolutely summable, we can freely swap the order of the sums and any limits herein. So using this and Lemma 13 gives us

$$
\begin{aligned}
v(f,P) &= \lim_{N\to\infty} \frac{1}{N}\mathbb{V}\mathrm{ar}(\sum_{i=1}^{N} f(X_i)) \\
&= \lim_{N\to\infty}[\sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{\infty}\sum_{i=1}^{n-1}\mathbb{I}_{k\le N-1}(k)\frac{N-k}{N}a_i^2\lambda_i^k] \\
&= \sum_{i=1}^{n-1}(a_i^2) + 2\lim_{N\to\infty}[\sum_{k=1}^{\infty}\sum_{i=1}^{n-1}\mathbb{I}_{k\le N-1}(k)\frac{N-k}{N}a_i^2\lambda_i^k] \\
&= \sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{\infty}\sum_{i=1}^{n-1}\lim_{N\to\infty}[\mathbb{I}_{k\le N-1}(k)\frac{N-k}{N}a_i^2\lambda_i^k] \\
&= \sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{\infty}\sum_{i=1}^{n-1}a_i^2\lambda_i^k \\
&= \sum_{i=1}^{n-1}(a_i^2) + 2\sum_{i=1}^{n-1}\sum_{k=1}^{\infty}a_i^2\lambda_i^k \\
&= \sum_{i=1}^{n-1}(a_i^2) + 2\sum_{i=1}^{n-1}a_i^2\frac{\lambda_i}{1-\lambda_i} \\
&= \sum_{i=1}^{n-1}[(a_i^2)(\frac{1-\lambda_i}{1-\lambda_i} + 2\frac{\lambda_i}{1-\lambda_i})] \\
&= \sum_{i=1}^{n-1}(a_i^2)\frac{1+\lambda_i}{1-\lambda_i}.
\end{aligned}
$$

Now assume P is periodic. Notice that if $\lambda_* < 1$, then by the above work, the result follows.
By Lemma 62 in the appendix, as P is irreducible and reversible, $\lambda_i < 1$, $\forall i \in \{1,\dots,n-1\}$.
Thus the only case we need to consider, is when $\lambda_i = -1$ for some $i \in \{1,\dots,n-1\}$. We assume $\lambda_i = -1$ for $i \in \{l, l+1,\dots,n-1\} \subseteq \{1,\dots,n-1\}$.
Note that in (1), we can seperate out the $\{l,\dots,n-1\}$ terms to get

$$
\frac{1}{N}\mathbb{V}\mathrm{ar}(\sum_{i=1}^{N} f(X_i)) = \sum_{i=0}^{n-1}(a_i^2)+2\sum_{k=1}^{N}\frac{N-k}{N}\sum_{i=0}^{l-1}a_i^2\lambda_i^k+2\sum_{i=l}^{n-1}(a_i^2)\sum_{k=1}^{N}\frac{N-k}{N}(-1)^k.
$$

Then in $\sum_{k=1}^{N} \frac{N-k}{N}(-1)^k$, replacing k with $2m-1$ if k is even and $2m-1$ if k is odd, and setting $E = \{k \in \mathbb{N} : k \text{ is even}\}$ and $O = \{k \in \mathbb{N} : k \text{ is odd}\}$,

$$\begin{aligned}
\sum_{k=1}^{N} \frac{N-k}{N}(-1)^k &= (\frac{1}{N}) \sum_{k=1}^{N} (N-k)(-1)^k \\
&= (\frac{1}{N}) \sum_{k=1}^{N} [\mathbb{I}_E(k)(N-k) + \mathbb{I}_O(k)(-N+k)] \\
&= (\frac{1}{N}) \sum_{m=1}^{\lfloor N/2 \rfloor} [(N-2m) + (-N+2m-1)] \\
&\text{(as if } N \text{ is odd, then the last term is } -N+N=0) \\
&= (\frac{1}{N}) \sum_{m=1}^{\lfloor N/2 \rfloor} -1 \\
&= \frac{\lfloor -N/2 \rfloor}{N}.
\end{aligned}$$

Note that in *Efficiency of Reversible MCMC Methods: Elementary Derivations and Applications to Composite Methods* by *Jeffrey Rosenthal* and *Radford Neal*, they write $\sum_{k=1}^{N} \frac{N-k}{N}(-1)^k = -\frac{\lfloor (N-1)/2 \rfloor}{N} - \frac{1}{N}\mathbb{I}_O(N-1)$, though we find here that this can be simplified to the above, $\frac{\lfloor -N/2 \rfloor}{N}$.

Now notice that

$$-1/2 = \lim_{k\to\infty} \frac{(N-1)/2}{N} \leq \lim_{k\to\infty} \frac{\lfloor N/2 \rfloor}{N} \leq \lim_{k\to\infty} \frac{N/2}{N} = -1/2,$$

so $\lim_{k\to\infty} \frac{\lfloor N/2 \rfloor}{N} = -1/2$, thus for $i \in \{l, \ldots, n-1\}$,

$$\lim_{k\to\infty} \sum_{k=1}^{N} \frac{N-k}{N}(-1)^k = -1/2 = \frac{-1}{1-(-1)} = \frac{\lambda_i}{1-\lambda_i}.$$

So, using this and the work done earlier when $\lambda_i < 1$, we get

$$\begin{aligned}
v(f,P) &= \lim_{N\to\infty}[\sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{N}\frac{N-k}{N}\sum_{i=1}^{l}a_i^2\lambda_i^k + 2\sum_{k=1}^{N}\frac{N-k}{N}\sum_{i=l}^{n-1}(a_i^2)\lambda_i^k] \\
&= \lim_{N\to\infty}[\sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{N}\frac{N-k}{N}\sum_{i=1}^{l}a_i^2\lambda_i^k + 2\sum_{k=1}^{N}\sum_{i=l}^{n-1}(a_i^2)\frac{N-k}{N}(-1)^k] \\
&= \lim_{N\to\infty}[\sum_{i=1}^{n-1}(a_i^2) + 2\sum_{k=1}^{N}\frac{N-k}{N}\sum_{i=1}^{l}a_i^2\lambda_i^k + 2\sum_{i=l}^{n-1}(a_i^2)\sum_{k=1}^{N}\frac{N-k}{N}(-1)^k] \\
&= \sum_{i=1}^{n-1}(a_i^2) + 2\sum_{i=1}^{n-2}a_i^2\frac{\lambda_i}{1-\lambda_i} + 2\sum_{i=l}^{n-1}a_i^2\frac{\lambda_i}{1-\lambda_i} \\
&= \sum_{i=1}^{n-1}(a_i^2)\frac{1+\lambda_i}{1-\lambda_i}.
\end{aligned}$$

□

Lemma 16. *If $h : \mathbb{R} \to \mathbb{R}$, and A is hermitian wrt $\langle\cdot,\cdot\rangle$ with real orthonormal basis of eigenvectors $\{v_0,\dots,v_{n-1}\}$, then $h(A) := \sum_{i=0}^{n-1} h(\lambda_i)v_iv_i^TD$ is also hermitian, and has the same real orthonormal basis of eigenvectors with associated eigenvalues $h(\lambda_0),\dots h(\lambda_{n-1})$.*

Proof. Let $h : \mathbb{R} \to \mathbb{R}$.
Let $\lambda_0,\dots,\lambda_{n-1}$ be the real eigenvalues of A and let $\{v_0,\dots,v_{n-1}\}$ be the assiciated orthonormal basis of real eigenvalues. Then as A is hermitian wrt $\langle\cdot,\cdot\rangle$, A has a spectral representation, i.e.

$$A = \sum_{i=0}^{n-1}\lambda_iv_iv_i^TD.$$

So, $\forall j \in \{0, \ldots, n-1\}$, notice

$$
\begin{aligned}
h(A)v_j &:= \sum_{i=0}^{n-1} h(\lambda_i) v_i v_i^T D v_j \\
&= \sum_{i=0}^{n-1} h(\lambda_i)(v_i(1)\langle v_i, v_j\rangle, \ldots, v_i(n)\langle v_i, v_j\rangle) \\
&= \sum_{i=0}^{n-1} h(\lambda_i)(v_i(1)\delta_{ij}, \ldots, v_i(n)\delta_{ij}) \\
&= h(\lambda_j)v_j.
\end{aligned}
$$

□

The above formula for the asymptotic variance (Proposition 15) gives us our first key theorem about Efficiency-Dominance.

Theorem 17. *If P is irreducible and reversible wrt π, then $\forall f \in L_0^2(\pi)$,*

$$v(f, P) = \langle f, f\rangle + 2\langle f, P(I-P)^{-1} f\rangle.$$

Proof. Let $f \in L_0^2(\pi)$. Then by Proposition 15, we know that

$$v(f, P) = \sum_{i=0}^{n-1} (a_i)^2 + 2\sum_{i=0}^{n-1} (a_i)^2 \frac{\lambda_i}{1-\lambda_i},$$

where $f = \sum_{i=0}^{n-1} a_i v_i$, and $\{v_0, \ldots, v_{n-1}\}$ is the real orthonormal basis of eigenvectors for P.

As $f \in L_0^2(\pi)$, we know that $0 = \mathbb{E}_\pi(f) = \langle v_0, f\rangle = \sum_{i=0}^{n-1} a_i \langle v_0, v_i\rangle = \sum_{i=0}^{n-1} a_i \delta_{0i} = a_0$, so we can restrict the rest of this to the subspace $W = \mathrm{span}(\{v_1, \ldots, v_{n-1}\})$.

Let $h : \mathbb{R}\backslash\{1\} \to \mathbb{R}$ such that $h(x) = x/(1-x)$.

Then as $\forall i \in \{1, \ldots, n-1\}$, $\lambda_i \neq 1$, on the restricted subspace W, $h(P)|_W =$

$P(I-P)^{-1}|_W$. So,

$$\begin{aligned}
\langle f, f\rangle + 2\langle f, P(I-P)^{-1}f\rangle &= \sum_{i=0}^{n-1}(a_i)^2 + 2\langle\sum_{j=0}^{n-1} a_j v_j, P(I-P)^{-1}(\sum_{i=0}^{n-1} a_i v_i)\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=0}^{n-1} a_i\langle v_j, P(I-P)^{-1}v_i\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=1}^{n-1} a_i\langle v_j, P(I-P)^{-1}|_W v_i\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=1}^{n-1} a_i\langle v_j, h(P)v_i\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=1}^{n-1} a_i\langle v_j, h(\lambda_i)v_i\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=1}^{n-1} a_i\langle v_j, \lambda_i(1-\lambda_i)^{-1}v_i\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=1}^{n-1} a_i\lambda_i(1-\lambda_i)^{-1}\langle v_j, v_i\rangle \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j \sum_{i=1}^{n-1} a_i\lambda_i(1-\lambda_i)^{-1}\delta_{ji} \\
&= \sum_{i=0}^{n-1}(a_i)^2 + 2\sum_{j=0}^{n-1} a_j^2\lambda_j(1-\lambda_j)^{-1}.
\end{aligned}$$

So by Proposition 15, $\langle f, f\rangle + 2\langle f, P(I-P)^{-1}f\rangle = v(f, P)$. □

This gives us our first conditions for efficiency dominance.

Corollary 18. *Given irreducible and reversible Markov Chains P and Q wrt π, P efficiency-dominates Q iff $\langle f, P(I-P)^{-1}f\rangle \leq \langle f, Q(I-Q)^{-1}f\rangle$ for all $f \in L_0^2(\pi)$.*

Remark. *Note that this result is actually stronger, that for each $f \in L_0^2(\pi)$, $v(f, P) \leq v(f, Q)$ iff $\langle f, P(I-P)^{-1}f\rangle \leq \langle f, Q(I-Q)^{-1}f\rangle$.*

Proof. By Theorem 17, $v(f,P) = \langle f,f\rangle + 2\langle f, P(I-P)^{-1}f\rangle$ for each $f \in L_0^2(\pi)$, and similarly for $v(f,Q)$.
Then $\forall f \in L_0^2(\pi)$, $v(f,Q) - v(f,P) \geq 0$ iff $\langle f, Q(I-Q)^{-1}f\rangle - \langle f, P(I-P)^{-1}f\rangle \geq 0$ iff $\langle f, Q(I-Q)^{-1}f\rangle \geq \langle f, P(I-P)^{-1}f\rangle$. □

Lemma 19. *Given linear transformations X, Y, and Z on any finite vector space V and $F : V \times V \to \mathbb{R}$, if Z is hermitian wrt F and $F(v, Xv) \leq F(v, Yv)$ for all $v \in V$, then $F(v, ZXZv) \leq F(v, ZYZv)$ for all $v \in V$.*

Proof. Notice that as Z is linear and V is a finite vector space, for any $v \in V$, $w = Zv \in V$. So, for any $v \in V$,

$$F(v, ZXZv) = F(Zv, XZv) = F(w, Xw) \leq F(w, Yw) = F(Zv, YZv) = F(v, ZYZv).$$

□

Definition. *A hermitian matrix A is strictly positive if $\langle v, Av\rangle > 0$ for every non-zero vector $v \in V$.*

Remark. *Note that because we are assuming the matrix A to be hermitian, there exists an orthormal basis of real eigenvectors $\{v_0, \ldots, v_{n-1}\}$ of A, with associated real eigenvalues $\{\lambda_0, \ldots, \lambda_{n-1}\}$. So for any $v \in V$, $\exists a_0, \ldots, a_{n-1} \in \mathbb{R}$ such that $v = \sum_{i=0}^{n-1} a_i v_i$. So, $\langle v, Av\rangle = \sum_{i=0}^{n-1}(a_i)^2\lambda_i$, and thus if A is a hermitian matrix, then A is strictly positive iff every eigenvalue of A is strictly positive.*

Lemma 20. *Given any $\alpha \in \mathbb{R}$, if A is a symmetric matrix wrt $\langle \cdot, \cdot \rangle$ with real eigenvalues and eigenvectors on the finite vector space V, then $\langle v, Av\rangle \leq \alpha\langle v, v\rangle$ for every $v \in V$ iff every eigenvalue λ of A is less than or equal to α. Similarly for $\geq$.*

Proof. Let $\alpha \in \mathbb{R}$ and $v \in V$. Let $\{w_0, \ldots, w_{n-1}\}$ be the basis of real orthonormal eigenvectors of A with associated eigenvalues $\lambda_0, \ldots, \lambda_{n-1} \in \mathbb{R}$.
Then $\exists a_0, \ldots, a_{n-1} \in \mathbb{R}$ such that $v = \sum_{i=0}^{n-1} a_i w_i$.
So, $\langle v, Av\rangle = \sum_{i=0}^{n-1}(a_i)^2\lambda_i$. So, if all the $\lambda_i \leq \alpha$, then $\sum_{i=0}^{n-1}(a_i)^2\lambda_i \leq \alpha\sum_{i=0}^{n-1}(a_i)^2 = \alpha\langle v, v\rangle$.
And if $\langle v, Av\rangle \leq \alpha\langle v, v\rangle$ for every $v \in V$, then for each $i \in \{0, \ldots, n-1\}$, picking the eigenvector w_i,

$$\lambda_i\langle w_i, w_i\rangle = \langle w_i, \lambda_i w_i\rangle = \langle w_i, Aw_i\rangle \leq \alpha\langle w_i, w_i\rangle,$$

so $\lambda_i \leq \alpha$ (as w_i is orthonormal wrt $\langle \cdot, \cdot \rangle$). □

Lemma 21. *If A and B are strictly positive hermitian matrices on the finite vector space V, then $\langle v, Av\rangle \leq \langle v, Bv\rangle$ for every $v \in V$ iff $\langle v, A^{-1}v\rangle \geq \langle v, B^{-1}v\rangle$ for every $v \in V$.*

Proof. Let A and B be strictly positive hermitian matrices on V. Assume $\langle v, Av\rangle \leq \langle v, Bv\rangle$ for every $v \in V$.
As A and B are both strictly positive, all the eigenvalues of A and B are in $(0, \infty)$ by the earlier remark.
So, for every eigenvalue β of B, $h(\beta) = 1/\sqrt{\beta}$ is defined for $h : (0, \infty) \to \mathbb{R}$ such that $h(x) = 1/\sqrt{x}$ for all $x \in (0, \infty)$. So, by Lemma 16, $B^{-1/2} = h(B)$ is hermitian wrt $\langle\cdot,\cdot\rangle$, and as each $\beta > 0$, $1/\sqrt{\beta} = h(\beta) > 0$, so $B^{-1/2}$ is also strictly positive.
To show that $B^{-1/2}AB^{-1/2}$ is also strictly positive, notice for any non-zero $v \in V$,

$$\begin{aligned}\langle v, B^{-1/2}AB^{-1/2}v\rangle &= \langle B^{-1/2}v, AB^{-1/2}v\rangle && \text{(as } B^{-1/2} \text{ is hermitian)}\\ &= \langle w, Aw\rangle \\ &\qquad \text{(where } w = B^{-1/2}v\text{, notice } w \neq 0 \text{ as } B^{-1/2} \text{ is strictly positive)}\\ &> 0. && \text{(as } w \text{ is non-zero and } A \text{ is strictly positive)}\end{aligned}$$

Now, as $B^{-1/2}$ is hermitian wrt $\langle\cdot,\cdot\rangle$, $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$, choosing $X = A$, $Y = B$ and $Z = B^{-1/2}$ in Lemma 19, by our assumption we get for every $v \in V$,

$$\langle v, B^{-1/2}AB^{-1/2}v\rangle \leq \langle v, B^{-1/2}BB^{-1/2}v\rangle = \langle v, v\rangle.$$

So by Lemma 20, the eigenvalues of $B^{-1/2}AB^{-1/2}$, say $\lambda_0, \dots, \lambda_{n-1} \in \mathbb{R}$, are less than or equal to 1.
And as $B^{-1/2}AB^{-1/2}$ is strictly positive, $\lambda_i > 0$ for every $i \in \{0, \dots, n-1\}$.
So, $\forall i \in \{0, \dots, n-1\}$, $\lambda_i \in (0, 1]$.
So, letting $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $g(x) = 1/x$ for every $x \in \mathbb{R}\backslash\{0\}$, as $\{\lambda_0, \dots, \lambda_{n-1}\} \subseteq (0, 1] \subseteq \mathbb{R}\backslash\{0\}$, by Lemma 16, $g(B^{-1/2}AB^{-1/2}) = (B^{-1/2}AB^{-1/2})^{-1}$ has all eigenvalues $g(\lambda_0), \dots, g(\lambda_{n-1}) \in [1, \infty)$.
So again by Lemma 19, we get that for every $v \in V$, $\langle v, (B^{-1/2}AB^{-1/2})^{-1}v\rangle \geq \langle v, v\rangle$. So,

$$\langle v, Iv\rangle = \langle v, v\rangle \leq \langle v, (B^{-1/2}AB^{-1/2})^{-1}v\rangle = \langle v, B^{1/2}A^{-1}B^{1/2}v\rangle.$$

Now again by Lemma 19, using $X = I$, $Y = B^{1/2}A^{-1}B^{1/2}$ and $Z = B^{-1/2}$, for every $v \in V$,

$$\langle v, B^{-1}v\rangle = \langle v, B^{-1/2}IB^{-1/2}v\rangle \leq \langle v, B^{-1/2}(B^{1/2}A^{-1}B^{1/2})B^{-1/2}v\rangle = \langle v, A^{-1}v\rangle.$$

For the other direction, note that from earlier using $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $g(x) = x^{-1}$ for every $x \in \mathbb{R}\backslash\{0\}$ on A and B create by Lemma 16 $h(A) = A^{-1}$ and $h(B) = B^{-1}$ that are strictly positive and hermitian on V. Thus it follows by replacing A with B^{-1} and B with A^{-1} from the start of this proof. $\square$

Proposition 22. *If P and Q are Markov Chains that are irreducible and reversible wrt π, then $\forall f \in L_0^2(\pi)$, $\langle f, P(I-P)^{-1}f\rangle \leq \langle f, Q^(I-Q)^{-1}f\rangle$ iff $\forall f \in L_0^2(\pi)$, $\langle f, Pf\rangle \leq \langle f, Qf\rangle$.*

Proof. Note that $\langle f, Pf\rangle \leq \langle f, Qf\rangle$ for all $f \in L_0^2(\pi)$ iff $\langle f, (I-P)f\rangle = \langle f, f\rangle - \langle f, Pf\rangle \geq \langle f, f\rangle - \langle f, Qf\rangle = \langle f, (I-Q)f\rangle$ for all $f \in L_0^2(\pi)$.
Next, notice that as $\forall f \in L_0^2(\pi)$, $0 = \mathbb{E}_\pi(f) = \langle f, 1\rangle$, and as 1 is an eigenvector of both P and Q with associated eigenvector $\lambda_0 = 1$, on the subspace $L_0^2(\pi)$, the eigenvalues of P and Q both exist in $[-1, 1)$.
So, using $h : \mathbb{R} \to \mathbb{R}$ such that $h(x) = 1 - x$ for every $x \in \mathbb{R}$, by Lemma 16 the eigenvalues of $h(P) = I - P$ and $h(Q) = I - Q$ are both contained in $(0, 2]$, so $I - P$ and $I - Q$ are both strictly positive and hermitian.
Thus, by the above Lemma 21, $\langle f, (I-P)^{-1}f\rangle \leq \langle f, (I-Q)^{-1}f\rangle$ for all $f \in L_0^2(\pi)$.
And as $(I-P)^{-1} = P(I-P)^{-1} + (I-P)(I-P)^{-1} = P(I-P)^{-1} + I$ and similarly for $(I-Q)^{-1}$, this is equivalent to

$$\langle f, P(I-P)^{-1}f\rangle \leq \langle f, Q(I-Q)^{-1}f\rangle, \quad \forall f \in L_0^2(\pi).$$

$\square$

We now get to our biggest theorems on efficiency dominance, which will serve as the backbone for the rest of the discussion on the topic.

Theorem 23. *Given irreducible reversible wrt π Markov Chains P and Q, P efficiency dominates Q iff $\langle f, Pf\rangle \leq \langle f, Qf\rangle$ for all $f \in L_0^2(\pi)$.*

Proof. By Corollary 18, we get that P efficiency dominates Q iff $\langle f, P(I-P)^{-1}f\rangle \leq \langle f, Q(I-Q)^{-1}f\rangle$ for all $f \in L_0^2(\pi)$.
By Proposition 22, this is iff $\langle f, Pf\rangle \leq \langle f, Qf\rangle$ for all $f \in L_0^2(\pi)$. $\square$

Theorem 24. *Given irreducible and reversible wrt π Markov Chains P and Q, P efficiency dominates Q iff $Q - P$ has all non-negative eigenvalues.*

Proof. By Theorem 23, P efficiency dominates Q iff $\langle f, Pf\rangle \leq \langle f, Qf\rangle$ for all $f \in L_0^2(\pi)$.
This is iff $\langle f, (Q-P)f\rangle \geq 0$ for every $f \in L_0^2(\pi)$.
So by Lemma 20, as $0 = 0\langle f, f\rangle$ for every $f \in L_0^2(\pi)$, the result follows. $\square$

Theorem 25. *Efficiency dominance is a partial order on irreducible reversible chains, meaning efficiency dominance is reflexive, antisymmetric and transitive.*

Proof. For any P and any $f : \mathcal{X} \to \mathbb{R}$, $v(f,P) \leq v(f,P)$, so it is reflexive.
If P and Q are irreducible reversible wrt π Markov Chains, then if P efficiency dominates Q and Q effieciency dominates P, then by Theorem 24, $Q-P$ and $P-Q$ have all nonnegative eigenvalues, and thus $Q-P$ must have all eigenvalues zero. So, as $Q-P$ is hermitian, it is diagonalizable, and thus $Q-P = A(\mathrm{diag}(0))A^{-1} = 0$ for some change of basis matrix A, so $Q = P$. Thus, it is antisymmetric.
As $\leq$ is transitive, we see that if P efficiency dominates Q and Q efficiency dominates R, then $\forall f \in L_0^2(\pi)$, $v(f,P) \leq v(f,Q)$ and $v(f,Q) \leq v(f,R)$, so by the transitivity of $\leq$, $v(f,P) \leq v(f,R)$. And thus P efficiency dominates R, and it is thus transitive. $\square$

6 New Results on Efficiency Dominance on Subspaces and Answer to Open Problem

All the results presented in this section are new. We first explore an explicit function in $L_0^2(\pi)$, f, such that $v(f,P) > v(f,Q)$ when $Q-P$ has a negative eigenvalue, as well as explore efficiency dominance on perpendicular subspaces of $L_0^2(\pi)$.

Lemma 26. $v_0 = (1, \ldots, 1)$ *is an eigenvector of* $Q-P$ *with eigenvalue* $\Lambda = 0$, *for any stochastic matrices* Q *and* P.

Proof.

$$\begin{aligned}
(Q-P)v_0 &= \begin{bmatrix} Q(0,0)-P(0,0) & \cdots & Q(0,n-1)-P(0,n-1) \\ \vdots & \ddots & \vdots \\ Q(n-1,0)-P(n-1,0) & \cdots & Q(n-1,n-1)-P(n-1,n-1) \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{x\in\mathcal{X}}[Q(0,x)-P(0,x)] \\ \vdots \\ \sum_{x\in\mathcal{X}}[Q(n-1,x)-P(n-1,x)] \end{bmatrix} \\
&= \begin{bmatrix} \sum_{x\in\mathcal{X}} Q(0,x) - \sum_{x\in\mathcal{X}} P(0,x) \\ \vdots \\ \sum_{x\in\mathcal{X}} Q(n-1,x) - \sum_{x\in\mathcal{X}} P(n-1,x) \end{bmatrix} \\
&= \begin{bmatrix} 1-1 \\ \vdots \\ 1-1 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \\
&= 0v_0.
\end{aligned}$$

□

Proposition 27. *If z is an eigenvector of $Q-P$ such that $(Q-P)z = \beta_0 z$ where $\beta_0 \in (-\infty, 0)$ and Q and P are both reversible wrt π and irreducible, then $z \in L_0^2(\pi)$.*

Proof. By Lemma 26, v_0 is an eigenvector of Q, P and $Q-P$, the last with eigenvalue 0. So, let $\{v_0, p_1, \ldots, p_{n-1}\}$ be a basis of real orthonormal eigenvectors of P, guaranteed to exist by Lemma 11. Similarly, let $\{v_0, q_1, \ldots, q_{n-1}\}$ be such a basis for Q.
Then we can write z as a linear combination of these vecotrs, so let $z = \Lambda v_0 + a_1 p_1 + \cdots + a_{n-1}p_{n-1} = \Lambda v_0 + b_1 q_1 + \cdots + b_{n-1}q_{n-1}$.
Then let ρ be the associated eigenvalue for z for P, and let η be the associated eigenvalue for z for Q. Then,

$$\rho\Lambda v_0 + \rho\sum_{i=1}^{n-1} a_i p_i = \rho z = P(z) = P(\Lambda v_0 + \sum_{i=1}^{n-1} a_i p_i) = \Lambda P(v_0) + P(\sum_{i=1}^{n-1} a_i p_i) = \Lambda v_0 + \sum_{i=1}^{n-1} \lambda_i a_i p_i.$$

As the vectors are linearly independent, either $\rho = 1$ or $\Lambda = 0$ and thus z is orthogonal to v_0 and $z \in L_0^2(\pi)$.
If $\rho = 1$, as P is diagonalizable by Lemma 11, either $z = v_0$ or the eigenspace of the eigenvalue $\lambda_0 = 1$ has dimension 2.
Notice that $(Q - P)(z) = \beta_0 z$ and $(Q - P)(v_0) = 0$ by Lemma 26, so as $\beta_0 \neq 0$, $(Q - P)(z) = \beta_0 z \neq 0 = (Q - P)(v_0)$. So $z \neq v_0$.
If the dimension of the eigenspace of the eigenvalue $\lambda_0 = 1$ has dimension 2, then $\lambda_1 = 1$. But by Lemma 37, $\lambda_1 < 1$. So another contradiction.
Thus $\rho \neq 1$, and $z \in L_0^2(\pi)$. □

Lemma 28. *If v is an eigenvector of $X, X - Y \in M_{n \times n}$ then z is an eigenvector of Y.*

Proof. Let $v \in V$ be an eigenvector of X with eigenvalue λ_X and of $X - Y$ with eigenvalue λ_{X-Y}. Then

$$\lambda_{X-Y} v = (X - Y)(v) = X(v) - Y(v) = \lambda_X v - Y(v).$$

So,

$$Y(v) = \lambda_X v - \lambda_{X-Y} v = (\lambda_X - \lambda_{X-Y})v.$$

□

In particular, this shows that if an eigenvector of $Q - P$, say z, is also an eigenvector of either Q or P, then it is also an eigenvector of both Q and P.

Using this fact and the remark after Lemma 21 in the efficiency dominance paper by Rosenthal and Neal, we provide a partial answer to the open problem left by Rosenthal and Neal.

Theorem 29. *If z is an eigenvector of $Q - P$ with associated eigenvalue $\beta < 0$ and z is also an eigenvector of P or Q, then $v(z, P) > v(z, Q)$.*

Proof. As z is an eigenvector of $Q - P$ and of either P or Q, by Lemma 28 it is an eigenvector of both Q and P.
By Proposition 27, $z \in L_0^2(\pi)$.
And as $(Q - P)(z) = \beta z$, $\langle z, (Q - P)(z)\rangle = \langle z, \beta z\rangle = \beta\langle z, z\rangle$. So, $\langle z, (Q - P)(z)\rangle < 0$, as z is not the zero vector as it is an eigenvector so $\langle z, z\rangle > 0$.
So, $\langle z, P(z)\rangle > \langle z, Q(z)\rangle$. Let $c = \langle z, z\rangle^{-1/2} > 0$. Then

$$\langle cz, P(cz)\rangle = c^2\langle z, P(z)\rangle > c^2\langle z, Q(z)\rangle = \langle cz, Q(cz)\rangle.$$

This is equivalent to $\langle cz, (I-P)(cz)\rangle = \langle cz, cz\rangle - \langle cz, P(cz)\rangle < \langle cz, cz\rangle - \langle cz, Q(cz)\rangle = \langle cz, (I-Q)(cz)\rangle$.
Further notice as $I - P = h(P)$ where $h : \mathbb{R} \to \mathbb{R}$ such that $h(x) = 1 - x$, by Lemma 16, cz is also an eigenvector of $h(P) = I - P$. Similarly cz is an eigenvector of Q.
By Lemma 62 in the appendix, if cz is an eigenvector of P and $cz \neq v_0 = (1, \dots, 1)$, then the associated eigenvalue of cz, $\lambda < 1$. Similary for Q, the associated eigenvalue of cz, $\alpha < 1$.
Thus, $h(\lambda), h(\alpha) < 0$, so $\langle cz, (I-P)(cz)\rangle = h(\lambda)\langle cz, cz\rangle \neq 0$. Similarly for $I - Q$.
Also, as $(I-P)^{-1} = u(I-P)$ where $u : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $u(x) = x^{-1}$, as the associated eigenvalue of cv wrt $I - P$ $h(\lambda) \neq 0$, by Lemma 16 cz is also an eigenvector of $(I-P)^{-1}$ with eigenvalue $u(h(\lambda)) = (h(\lambda))^{-1}$, and similarly for $I - Q$ with eigenvalue $u(h(\alpha)) = (h(\alpha))^{-1}$.
So, we get

$$\begin{aligned}
\langle cz, (I-P)^{-1}(cz)\rangle &= c^2\langle z, (I-P)^{-1}(z)\rangle \\
&= c^2\langle z, (h(\lambda))^{-1}z\rangle \\
&= (h(\lambda))^{-1}c^2\langle z, z\rangle \\
&= (h(\lambda))^{-1} \\
&= (h(\lambda)c^2\langle z, z\rangle)^{-1} \\
&= (\langle cz, (I-P)(cz)\rangle)^{-1} \\
&> (\langle cz, (I-Q)(cz)\rangle)^{-1} \\
&= (h(\alpha)c^2\langle z, z\rangle)^{-1} \\
&= (h(\alpha))^{-1} \\
&= (h(\alpha))^{-1}c^2\langle z, z\rangle \\
&= \langle cz, (I-Q)^{-1}(cz)\rangle.
\end{aligned}$$

So, $\langle cz, (I-P)^{-1}(cz)\rangle > \langle cz, (I-Q)^{-1}(cz)\rangle$.
Notice that $(I-P)^{-1} = P(I-P)^{-1} + (I-P)(I-P)^{-1} = P(I-P)^{-1} + I$, so this becomes

$$\begin{aligned}
\langle cz, P(I-P)^{-1}(cz)\rangle + \langle cz, cz\rangle &= \langle cz, (I-P)^{-1}(cz)\rangle \\
&> \langle cz, (I-Q)^{-1}(cz)\rangle \\
&= \langle cz, Q(I-Q)^{-1}(cz)\rangle + \langle cz, cz\rangle.
\end{aligned}$$

So, we get that

$$\langle cz, P(I-P)^{-1}(cz)\rangle > \langle cz, Q(I-Q)^{-1}(cz)\rangle.$$

Thus by Theorem 17, $v(cz,P) > v(cz,Q)$.
Further notice that

$$v(cz,P) = \lim_{N\to\infty}(1/N)\mathbb{V}\mathrm{ar}(\sum_{i=1}^{N} cz(X_i)) = c^2 \lim_{N\to\infty}(1/N)\mathbb{V}\mathrm{ar}(\sum_{i=1}^{N}(z(X_i)) = c^2 v(z,P).$$

Similarly, $v(cz,Q) = c^2 v(z,Q)$.
So, $c^2 v(z,P) = v(cz,P) > v(cz,Q) = c^2 v(z,Q)$, and as $c^2 > 0$,

$$v(z,P) > v(z,Q).$$

□

Proposition 30. *If $v \in N \subseteq L_0^2(\pi)$ where $N := span\{w_l, \ldots, w_{n-1}\}^{\perp}|_{L_0^2(\pi)} = \{v \in L_0^2(\pi) : \langle v, w\rangle, \forall w \in span\{w_l, \ldots, w_{n-1}\}\}$, then $Pv, Qv \in N$. I.e. N is closed under P and Q.*

Proof. Let $v \in N$. Then $\forall w \in \mathrm{span}\{w_l, \ldots, w_{n-1}\}$, $\exists \alpha_l, \ldots, \alpha_{n-1} \in \mathbb{R}$ such that $w = \alpha_l w_l + \cdots + \alpha_{n-1}w_{n-1}$. Let λ_i denote the eigenvalue of w_i wrt P. Then,

$$\begin{aligned}\langle P(v), w\rangle &= \langle v, P(w)\rangle \\ &= \langle v, P(\sum_{i=l}^{n-1}\alpha_i w_i)\rangle \\ &= \sum_{i=l}^{n-1}\alpha_i\langle v, P(w_i)\rangle \\ &= \sum_{i=l}^{n-1}\alpha_i\langle v, \lambda_i w_i\rangle \\ &= \sum_{i=l}^{n-1}\alpha_i\lambda_i\langle v, w_i\rangle \\ &= 0.\end{aligned}$$

Similarly for Q. □

We now generalize another few Lemmas, in order to try and apply them to a specific subspace of $L_0^2(\pi)$.

Lemma 31. *Given linear transormations X, Y, and Z on any finite vector space V, and $F : V \times V \to \mathbb{R}$, if Z is hermitian wrt F, $F(v, Xv) \leq F(v, Yv)$ for all $v \in E \subseteq V$ where E is a subspace of V, and E is closed wrt Z, then $F(v, ZXZv) \leq F(v, ZYZv)$ for all $v \in E$.*

Proof. As E is closed wrt Z, by definition for any $v \in E$, $Zv = w \in E$. So for any $v \in E$,

$$F(v, ZXZv) = F(Zv, XZv) = F(w, Xw) \leq F(w, Yw) = F(Zv, YZv) = F(v, ZYZ).$$

□

Lemma 32. *If A and B are strictly positive hermitian matrices on a subspace E of the vector space V, and E is closed wrt A and B, then $\langle v, Av\rangle \leq \langle v, Bv\rangle$ for every $v \in E$ iff $\langle v, A^{-1}v\rangle \geq \langle v, B^{-1}v\rangle$ for every $v \in E$.*

Proof. Let $E \subseteq V$ be a subspace of V and let A and B be strictly positive hermitian matrices on E such that E is closed wrt A and B. Assume $\langle v, Av\rangle \leq \langle v, Bv\rangle$ for every $v \in E$.
As A and B are both strictly positive, all the eigenvalues of A and B for eigenvectors in E are in $(0, \infty)$.So, for every eigenvalue on E, β of B, $h(\beta) = 1/\sqrt{\beta}$ is defined for $h : (0, \infty) \to \mathbb{R}$ such that $h(x) = 1/\sqrt{x}$ for all $x \in (0, \infty)$. So, by Lemma 16, $B^{-1/2} = h(B)$ is hermitian wrt $\langle \cdot, \cdot \rangle$.
As E is closed under B, matrix decomposition simply affects the scalar values of vectors multiplied, but not the geometric propoerties of vectors coming from B. So, E is also closed wrt $B^{-1/2}$. And as each $\beta > 0$, $1/\sqrt{\beta} = h(\beta) > 0$, so $B^{-1/2}$ is also strictly positive on E.
To show that $B^{-1/2}AB^{-1/2}$ is also strictly positive on E, notice for any non-zero $v \in E$,

$$\begin{aligned}
\langle v, B^{-1/2}AB^{-1/2}v\rangle &= \langle B^{-1/2}v, AB^{-1/2}v\rangle \\
&\qquad \text{(as } B^{-1/2} \text{ is hermitian)} \\
&= \langle w, Aw\rangle \\
&\qquad \text{(where } w = B^{-1/2}v\text{, notice } w \neq 0 \text{ as } B^{-1/2} \text{ is strictly positive on } E\text{)} \\
&> 0. \\
&\qquad \text{(as } w \text{ is non-zero and } A \text{ is strictly positive on } E\text{)}
\end{aligned}$$

Now, as $B^{-1/2}$ is hermitian wrt $\langle\cdot,\cdot\rangle$, $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$, choosing $X = A$, $Y = B$ and $Z = B^{-1/2}$ in Lemma 31, by our assumption we get for every $v \in E$,

$$\langle v, B^{-1/2}AB^{-1/2}v\rangle \leq \langle v, B^{-1/2}BB^{-1/2}v\rangle = \langle v, v\rangle.$$

So by Lemma 20, the eigenvalues of $B^{-1/2}AB^{-1/2}$ on the restricted subspace E, say $\lambda_0, \dots, \lambda_{k-1} \in \mathbb{R}$, are less than or equal to 1.
And as $B^{-1/2}AB^{-1/2}$ is strictly positive, $\lambda_i > 0$ for every $i \in \{0, \dots, k-1\}$.
So, $\forall i \in \{0, \dots, k-1\}$, $\lambda_i \in (0, 1]$.
So, letting $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $g(x) = 1/x$ for every $x \in \mathbb{R}\backslash\{0\}$, as $\{\lambda_0, \dots, \lambda_{k-1}\} \subseteq (0, 1] \subseteq \mathbb{R}\backslash\{0\}$, by Lemma 16, $g(B^{-1/2}AB^{-1/2}) = (B^{-1/2}AB^{-1/2})^{-1}$ has all eigenvalues $g(\lambda_0), \dots, g(\lambda_{k-1}) \in [1, \infty)$.
So again by Lemma 31, we get that for every $v \in E$, $\langle v, (B^{-1/2}AB^{-1/2})^{-1}v\rangle \geq \langle v, v\rangle$. So,

$$\langle v, Iv\rangle = \langle v, v\rangle \leq \langle v, (B^{-1/2}AB^{-1/2})^{-1}v\rangle = \langle v, B^{1/2}A^{-1}B^{1/2}v\rangle.$$

Now again by Lemma 31, using $X = I$, $Y = B^{1/2}A^{-1}B^{1/2}$ and $Z = B^{-1/2}$, for every $v \in E$,

$$\langle v, B^{-1}v\rangle = \langle v, B^{-1/2}IB^{-1/2}v\rangle \leq \langle v, B^{-1/2}(B^{1/2}A^{-1}B^{1/2})B^{-1/2}v\rangle = \langle v, A^{-1}v\rangle.$$

For the other direction, note that from earlier using $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $g(x) = x^{-1}$ for every $x \in \mathbb{R}\backslash\{0\}$ on A and B create by Lemma 16 $h(A) = A^{-1}$ and $h(B) = B^{-1}$ that are strictly positive and hermitian on E, and E is still closed wrt A^{-1} and B^{-1}. Thus the result follows by replacing A with B^{-1} and B with A^{-1} from the start of the proof. $\square$

Proposition 33. *If P and Q are Markov Chains that are irreducible and reversible wrt π and $E \subseteq L_0^2(\pi)$ is a subspace that is closed wrt P and Q, then $\forall f \in E$, $\langle f, P(I-P)^{-1}f\rangle \leq \langle f, Q(I-Q)^{-1}f\rangle$ iff $\forall f \in E$, $\langle f, Pf\rangle \leq \langle f, Qf\rangle$.*

Proof. Note that $\langle f, Pf\rangle \leq \langle f, Qf\rangle$ for all $f \in E$ iff $\langle f, (I-P)f\rangle = \langle f, f\rangle - \langle f, Pf\rangle \geq \langle f, f\rangle - \langle f, Qf\rangle = \langle f, (I-Q)f\rangle$ for all $f \in E$.
Next, notice that as $\forall f \in E \subseteq L_0^2(\pi)$, $0 = \mathbb{E}_\pi(f) = \langle f, 1\rangle$, and as 1 is an eigenvector of both P and Q with associated eigenvector $\lambda_0 = 1$, on the subspace $L_0^2(\pi)$, the eigenvalues of P and Q both exist in $[-1, 1)$, as by Lemma 62 in the appendix, they are all less than one.
So, using $h : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $h(x) = 1 - x$ for every $x \in \mathbb{R}\backslash\{0\}$, by Lemma 16 the eigenvalues of $h(P) = I - P$ and $h(Q) = I - Q$ on E are

both contained in $(0, 2]$, so $I - P$ and $I - Q$ are both strictly positive and hermitian on E.
Thus, as E is closed wrt P, Q, and I (and thus also $I - P$ and $(I - P)^{-1}$ and similarly for Q), by the above Lemma 32, $\langle f, (I - P)^{-1} f\rangle \leq \langle f, (I - Q)^{-1} f\rangle$ for all $f \in E$.
And as $(I - P)^{-1} = P(I - P)^{-1} + (I - P)(I - P)^{-1} = P(I - P)^{-1} + I$ and similarly for $(I - Q)^{-1}$, this is equivalent to

$$\langle f, P(I - P)^{-1} f\rangle \leq \langle f, Q(I - Q)^{-1} f\rangle, \quad \forall f \in E.$$

$\square$

We now prove why $\text{span}\{w_l, \ldots, w_{n-1}\}^{\perp}|_{L_0^2(\pi)}$ is a special subspace.

Proposition 34. *If P and Q are irreducible and reversible wrt π Markov Chains, all the eigenvalues of $Q - P$, $\beta_0, \ldots, \beta_{n-1}$, that are negative, say $\beta_l, \ldots, \beta_{n-1}$, have associated eigenvalues $w_l, \ldots, w_{n-1}$ that are also eigenvalues of either Q or P, then $\forall v \in N \subseteq L_0^2(\pi)$ such that $N := \text{span}\{w_l, \ldots, w_{n-1}\}^{\perp}|_{L_0^2(\pi)} = \{v \in L_0^2(\pi) : \langle v, w\rangle, \forall w \in \text{span}\{w_l, \ldots, w_{n-1}\}\}$, $\langle v, Pv\rangle \leq \langle v, Qv\rangle$.*

Proof. Let $v \in N$. Then as Q and P are hermitian, so is $Q - P$. So by an argument similar to why Q and P have real eigenvalues and an orthonormal basis, so does $Q - P$ on $L_0^2(\pi)$.
By assumption, the $w_l, \ldots, w_{n-1}$ are also eigenvectors of either Q or P, so by Lemma 28, they are also an eigenvector of both P and Q. Thus we can assume they are orthonormal.
Then by a process like the Gram-Schmidt process, we can find other eigenvectors $w_0, \ldots, w_{l-1}$ that complete the orthonormal basis for $Q - P$.
So, $\exists \alpha_0, \ldots, \alpha_{n-1} \in \mathbb{R}$ such that $v = \sum_{i=0}^{n-1} \alpha_i w_i$.
But as $v \in N$, $\alpha_l, \ldots, \alpha_{n-1} = 0$. So,

$$\langle v, (Q - P)(v)\rangle = \sum_{i=0}^{l-1} \alpha_i^2 \beta_i.$$

And as $\alpha_i^2 \geq 0$ for every i, and $\beta_i \geq 0$ for every $i \in \{0, \ldots, l - 1\}$, $\langle v, (Q - P)(v)\rangle = \sum_{i=0}^{l-1} \alpha_i^2 \beta_i \geq 0$.
So, $\langle v, P(v)\rangle \leq \langle v, Q(v)\rangle$. $\square$

Theorem 35. *If P and Q are irreducible reversible wrt π Markov Chains and all the eigenvalues of $Q - P$, $\beta_0, \ldots, \beta_{n-1}$, that are negative, say $\beta_l, \ldots, \beta_{n-1}$,* *have associated eigenvalues $w_l, \ldots, w_{n-1}$ that are also eigenvalues of either Q or P, then P efficiency dominates Q on the subspace $N \subseteq L_0^2(\pi)$ such that $N := span\{w_l, \ldots, w_{n-1}\}^{\perp}|_{L_0^2(\pi)} = \{v \in L_0^2(\pi) : \langle v, w\rangle, \forall w \in span\{w_l, \ldots, w_{n-1}\}\}$.*

Proof. We know that P efficiency dominates Q on N if $\langle v, P(I-P)^{-1}v\rangle \leq \langle v, Q(I-Q)^{-1}v\rangle$ for every $v \in N$.
We know that N is closed wrt P and Q and thus also $Q - P$ by Proposition 30.
By Proposition 34, we know that $\langle v, Pv\rangle \leq \langle v, Qv\rangle$ for every $v \in N$.
So, by Proposition 33 with $E = N$, we know that $\langle v, P(I-P)^{-1}v\rangle \leq \langle v, Q(I-Q)^{-1}v\rangle$ for every $v \in N$, completing the proof. □

7 Combinations of Chains

We now move to using the results in the earlier section to apply to combinations of chains.

Proposition 36. *If P is an irreducible Markov Chain, and Q is another Markov chain, then for any $\alpha \in (0, 1]$, $P' = \alpha P + (1-\alpha)Q$ is an irreducible Markov Chain.*

Proof. Let $x \in \mathcal{X}$. Let $A \subseteq \mathcal{X}$. Then as P is irreducible, $\exists r \in \mathbb{N}$ such that $P^r(x, A) > 0$.
So, $P'^r(x, A) \geq \alpha^r P^r(x, A) > 0$. I.e., the probability that P' gets to A from x in r steps is at least the probability that we get P r times, and get to A the same way P can. □

Theorem 37. *If P, P' and Q are reversible wrt π Markov Chains such that P and P' are irreducible, if $\alpha \in (0, 1]$, then P' efficiency dominates P iff $\alpha P' + (1-\alpha)Q$ efficiency dominates $\alpha P + (1-\alpha)Q$.*

Proof. By the above Proposition 36, both $\alpha P' + (1-\alpha)Q$ and $\alpha P + (1-\alpha)Q$ are irreducible.
As P' efficiency dominates P iff the eigenvalues of $P - P'$ are all non-negative, by combining Theorem 24, P' efficiency dominates P iff $\alpha(P - P')$ has all eigenvalues non-negative as $\alpha > 0$. So, as

$$\alpha(P - P') = (\alpha P + (1-\alpha)Q) - (\alpha P' + (1-\alpha)Q),$$

the result follows. □

Theorem 38. *For any $l \in \mathbb{N}$, let $P_1, \dots, P_l, P'_1, \dots, P'_l$ be reversible Markov Chains wrt π. Assume $P = \alpha_1 P_1 + \cdots + \alpha_l P_l$ and $P' = \alpha_1 P'_1 + \cdots + \alpha_l P'_l$ are irrducible Markov Chains, with $\alpha_1, \dots, \alpha_l > 0$ and $\sum_{i=1}^{l} \alpha_i = 1$. Then if $\forall i \in \{1, \dots, l\}$, the eigenvalues of $P_i - P'_i$ are non-negative, then P' efficiency dominates P.*

Proof. Let $f \in L_0^2(\pi)$. Then by Lemma 20, $\forall i \in \{1, \dots, l\}$, $\langle f, (P_i - P'_i) f \rangle \geq 0 \langle f, f \rangle = 0$, so

$$\langle f, (P - P')f \rangle = \langle f, \sum_{i=1}^{l} (P_i - P'_i) f \rangle = \sum_{i=1}^{l} \langle f, (P_i - P'_i) f \rangle \geq 0.$$

So equivalently, $\langle f, P' f \rangle \leq \langle f, P f \rangle$. So by Theorem 23, P' efficiency dominates P. $\square$

8 Not Dominatable Chains

We now present results showing chains that don't dominate each other.

Definition (Eigen Dominance)**.** *Given two reversible wrt π Markov Chains P and Q, and their eigenvalues written in non-increasing order, $\alpha_0, \dots, \alpha_{n-1}$ for P and $\beta_0, \dots, \beta_{n-1}$ for Q, P is said to eigen dominate Q if $\forall i \in \{0, \dots, n-1\}$, $\alpha_i \leq \beta_i$.*

We shall now see that with irreducible Markov Chains, efficiency dominance implies eigen dominance.

Proposition 39. *If P and Q are irreducible, reversible wrt π Markov Chains, such that P efficiency dominates Q, then P eigen-dominates Q.*

Proof. By Theorem 23, as P efficiency dominates Q, $\forall f \in L_0^2(\pi)$, $\langle f, P f \rangle \leq \langle f, Q f \rangle$.
So, by the "min-max" characterisation of eigenvalues, or the Courant-Fischer Theorem (Theorem 4.2.6) in *Matrix Analysis* by Horn and Johnson, we see that $\forall i \in \{0, \dots, n-1\}$, letting $\alpha_0, \dots, \alpha_{n-1}$ be the eigenvalues of P and $\beta_0, \dots, \beta_{n-1}$ be the the eigenvalues of Q,

$$\alpha_i = \min_{w_1, \dots, w_{i-1}} [\max_{v \in \mathbb{R}^{\mathcal{X}} : \forall j, \langle v, w_j \rangle = 0} \frac{\langle v, Pv \rangle}{\langle v, v \rangle}],$$

and similarly for β_i and Q.
So, assume for a contradiction that $\exists i \in \{0, \ldots, n-1\}$ such that $\alpha_i > \beta_i$.
Notice that $i \neq 0$, as $\alpha_i = \beta_i = 1$ by Proposition 3.
Then by the above,

$$\min_{w_1,\ldots,w_{i-1}} [\max_{v \in \mathbb{R}^{\mathcal{X}}:\forall j, \langle v, w_j \rangle = 0} \frac{\langle v, Pv \rangle}{\langle v, v \rangle}] > \min_{w_1,\ldots,w_{i-1}} [\max_{v \in \mathbb{R}^{\mathcal{X}}:\forall j, \langle v, w_j \rangle = 0} \frac{\langle v, Qv \rangle}{\langle v, v \rangle}].$$

So let $z \in \mathbb{R}^{\mathcal{X}}$ be the vector such that

$$\frac{\langle z, Qz \rangle}{\langle z, z \rangle} = \min_{w_1,\ldots,w_{i-1}} [\max_{v \in \mathbb{R}^{\mathcal{X}}:\forall j, \langle v, w_j \rangle = 0} \frac{\langle v, Qv \rangle}{\langle v, v \rangle}].$$

And as $i \neq 0$, we have $z \in L_0^2(\pi)$. So, we get

$$\frac{\langle z, Qz \rangle}{\langle z, z \rangle} \geq \frac{\langle z, Pz \rangle}{\langle z, z \rangle}.$$

But then

$$\min_{w_1,\ldots,w_{i-1}} [\max_{v \in \mathbb{R}^{\mathcal{X}}:\forall j, \langle v, w_j \rangle = 0} \frac{\langle v, Pv \rangle}{\langle v, v \rangle}] > \frac{\langle z, Pz \rangle}{\langle z, z \rangle}.$$

□

Note however the converse is not true. But, this lets us see that if P doesn't eigen-dominate Q, then P doesn't efficiency dominate Q.

Proposition 40. *Given irreducible reversible wrt π Markov Chains P and Q, if the eigenvalues of P and Q are identical, then $P \neq Q$ iff P doesn't efficiency dominate Q and Q doesn't efficiency dominate P.*

Proof. If $P = Q$, then as efficiency dominance is relfexive P efficiency dominates Q and Q efficiency dominates P.
If $P \neq Q$, then assume P efficiency dominates Q. Then by Theorem 24, $Q - P$ has all non-negative eigenvalues.
But $Q - P$ cannot have all zero eigenvalues, as then because $Q - P$ is hermitian and thus diagonalizable, $Q - P = A(\text{diag}(0))A^{-1} = 0$ for some change of basis matrix A, so $Q = P$.
So, there must be at least one positive eigenvalue of $Q - P$.
As the sum of eigenvalues of a matrix including multiplicity is equal to the trace of a matrix, we get that $\text{trace}(Q - P) > 0$.

But as trace is linear, $\text{trace}(Q-P) = \text{trace}(Q) - \text{trace}(P) > 0$, so $\text{trace}(Q) > \text{trace}(P)$.
But as P and Q have the same eigenvalues, $\text{trace}(Q) = \text{trace}(P)$. So we get a contradiction.
The argument for why Q doesn't efficiency dominate P follows by replacing Q with P and P with Q. $\square$

Lemma 41. *For any Markov Chain P with π as it's stationary distribution, trace$(P) \geq \max(0, 2 - 1/\pi_{\max})$ where $\pi_{\max} = \max_x \pi(x)$.*

Proof. If $\pi(x) \leq 1/2$ for all $x \in \mathcal{X}$, then $\text{trace}(P) \geq 0$ as each entry of P is non-negative.
If $\exists x_* \in \mathcal{X}$ such that $\pi(x_*) > 1/2$, then assume without loss of generality that $x_* = x_{n-1}$. Then

$$\begin{aligned}
\pi(x_{n-1}) &= \sum_{i=0}^{n-1} \pi(x_i) P(x_i, x_{n-1}) && \text{(as } \pi \text{ is a stationary dis.)} \\
&= \pi(x_{n-1}) P(x_{n-1}, x_{n-1}) + \sum_{i=0}^{n-2} \pi(x_i) P(x_i, x_{n-1}) \\
&\leq \pi(x_{n-1}) P(x_{n-1}, x_{n-1}) + \sum_{i=0}^{n-2} \pi(x_i) && \text{(as each } P(x_i, x_{n-1}) \leq 1.) \\
&= \pi(x_{n-1}) P(x_{n-1}, x_{n-1}) + (1 - \pi(x_{n-1})). && \Big(\text{as } \sum_{i=0}^{n-1} \pi(x_i) = 1\Big).
\end{aligned}$$

So, $P(x_{n-1}, x_{n-1}) \geq 2 - 1/\pi(x_{n-1}) = 2 - 1/\pi_{\max}$ (as if $\pi(x_{n-1}) > 1/2$, then no other x_i can have $\pi(x_i) > \pi(x_{n-1})$ as then $\sum_{j=0}^{n-1} \pi(x_j) \geq \pi(x_{n-1}) + \pi(x_i) > 1$).
So, $\text{trace}(P) \geq P(x_{n-1}, x_{n-1}) \geq 2 - 1/\pi_{\max}$. $\square$

Lemma 42. *If P is a Markov Chain with stationary distribution π and trace$(P) = \max(0, 2 - 1/\pi_{\max})$, then every entry on the diagonal is zero unless $\exists x \in \mathcal{X}$ such that $\pi(x) > 1/2$. In that case every diagonal entry is zero except for $P(x,x) = 2 - 1/\pi(x)$.*

Proof. We continue from the proof of the above Lemma 41. If $\pi(x) \leq 1/2$ for every $x \in \mathcal{X}$, then $\text{trace}(P) = 0$, and as all entries are non-negative all diagonal entries are zero.
If $x_{n-1} \in \mathcal{X}$ such that $\pi(x_{n-1}) > 1/2$, then $2 - 1/\pi(x_{n-1}) > 0$, so $\text{trace}(P) =$

$\max(0, 2 - 1/\pi(x_{n-1})) = 2 - 1/\pi(x_{n-1})$.
From the proof of the above Lemma 41, $P(x_{n-1}, x_{n-1}) \geq 2 - 1/\pi(x_{n-1})$. But notice

$$2 - 1/\pi(x) = \text{trace}(P) \geq P(x_{n-1}, x_{n-1}) \geq 2 - 1/\pi(x_{n-1}),$$

so $\text{trace}(P) = P(x_{n-1}, x_{n-1}) = 2 - 1/\pi(x_{n-1})$. $\square$

Theorem 43. *If P is an irreducible reversible wrt π Markov Chain and trace$(P) = \max(0, 2 - 1/\pi_{\max})$ where $\pi_{\max} = \max_{x \in \mathcal{X}} \pi(x)$, then P cannot be efficiency dominated by any other irreducible reversible wrt π Markov Chain.*

Proof. Assume for a contradiction that Q is an irreducible reversible wrt π Markov Chain that efficiency dominates P and $Q \neq P$.
Then by Proposition 39, Q also eigen-dominates P.
If Q has the same eigenvalues as P, then by Proposition 40, as $Q \neq P$, Q doesn't efficiency dominate P.
So, Q must have at least one eigenvalue strictly less than one of P. But then $\text{trace}(Q) < \text{trace}(P)$, as the trace is equal to the sum of the eigenvalues of the matrix. But by Lemma 41, $\text{trace}(Q) \geq \max(0, 2 - 1/\pi_{\max}$. So,

$$\max(0, 2 - 1/\pi_{\max}) \leq \text{trace}(Q) < \text{trace}(P) = \max(0, 2 - 1/\pi_{\max}).$$

So we get a contradiction. $\square$

9 Peskun versus Efficiency Dominance

We now move to show the relationship between Efficiency Dominance and Peskun Dominance.

Definition (Peskun Dominance)**.** *Given two Markov Chains, P Peskun-dominates Q if $\forall x, y \in \mathcal{X}$ such that $x \neq y$, $P(x, y) \geq Q(x, y)$.*

Now to show Peskun-dominance implies efficiency-dominance.

Lemma 44. *If P Peskun-dominates Q, then $Q - P$ has all non-negative values on the diagonal and all non-positive entries off the diagonal, and the sum of each row is 0. I.e. $\forall x, y \in \mathcal{X}$, if $x = y$ then $(Q - P)(x, y) \geq 0$ and if $x \neq y$ then $(Q - P)(x, y) \leq 0$, and $\forall x \in \mathcal{X}$, $\sum_{i=1}^{n} (Q - P)(x_i, x) = 0$.*

Proof. As P Peskun-dominates Q, by definition $\forall x, y \in \mathcal{X}$, $P(x,y) \geq Q(x,y)$ if $x \neq y$.
So, if $x, y \in \mathcal{X}$ such that $x \neq y$, then $(Q-P)(x,y) = Q(x,y) - P(x,y) \leq 0$.
Now, assume for a contradiction that $\exists x \in \mathcal{X}$ such that $(Q-P)(x,x) < 0$.
Then $0 > (Q-P)(x,x) = Q(x,x) - P(x,x)$, so $Q(x,x) < P(x,x)$.
So, as P Peskun-dominates Q, $\sum_{y \in \mathcal{X}, y \neq x} P(y,x) \geq \sum_{y \in \mathcal{X}, y \neq x} Q(y,x)$. So,

$$\sum_{y \in \mathcal{X}} P(y,x) = \sum_{y \in \mathcal{X}, y \neq x} P(y,x) + P(x,x) > \sum_{y \in \mathcal{X}, y \neq x} Q(y,x) + Q(x,x) = \sum_{y \in \mathcal{X}} Q(y,x) = 1.$$

So we get a contradiction to the Law of Total Probability. Let $x \in \mathcal{X}$. Then by the Law of Total Probability,

$$\sum_{i=0}^{n-1} (Q-P)(x_i, x) = \sum_{i=0}^{n-1} [Q(x_i,x) - P(x_i,x)] = \sum_{i=0}^{n-1} Q(x_i,x) - \sum_{i=0}^{n-1} P(x_i,x) = 1-10.$$

$\square$

Lemma 45. *If $X \in M^{n \times n}$ such that $\forall i, j \in \{0, \ldots, n-1\}$ and $i \neq j$, $X(i,i) \geq 0$ and $X(i,j) \leq 0$, and $\sum_{l=0}^{n-1} X(l,i) = 0$, then all the eigenvalues of X are non-negative.*

Proof. Let $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that $Xv = \lambda v$. Then let $j \in \{0, \ldots, n-1\}$ such that $|v_j| \geq |v_i|$ for all $i \in \{0, \ldots, n-1\}$.
Without loss of generality, assume $v_j > 0$. Then as $-\sum_{i \neq j} |X(j,i)| = \sum_{i \neq j} X(j,i)$ as they're all non-positive, as the sum of the rows of X is equal to 0, and by the triangle inequality,

$$\begin{aligned}
\lambda v_j = (Xv)_j = \sum_{i=0}^{n-1} X(j,i) v_j &= X(j,j) v_j + \sum_{i \neq j} X(j,i) v_i \\
&\geq X(j,j) v_j - |\sum_{i \neq j} X(j,i) v_i| \\
&\geq X(j,j) v_j - \sum_{i \neq j} |X(j,i)||v_i| \\
&\geq X(j,j) v_j - \sum_{i \neq j} |X(j,i)| v_j \\
&= v_j (X(j,j) + \sum_{i \neq j} X(j,i)) \\
&= v_j(0) = 0.
\end{aligned}$$

As $v_j > 0$, this means $\lambda \geq 0$. $\square$

Proposition 46. *If P and Q are irreducible reversible wrt π Markov Chains, and P Peskun-dominates Q, then P efficiency-dominates Q.*

Proof. By Lemma 44, $Q - P$ has all diagonal entries non-negative and all off diagonal entries non-positive, and has the sum of each row equal to zero.
So by Lemma 45, all the eigenvalues of $Q - P$ are non-negative, and thus by Theorem 24, P efficiency dominates Q. $\square$

10 Group state spaces and Random Walks

We move our discussion to a special type of Markov Chain on a special type of state space. The special properties these assumptions bring give us an easy, simple bound on the total variation distance of such Markov Chains.

So, we begin this chapter with the definition on the type of state space we will be working with.

Definition (Groups)**.** *A Group is a non-empty set G together with an operation, denoted $(G, \cdot)$, that satisfies the following.*

- $\exists 1 \in G$ *such that* $\forall x \in G$, $1 \cdot x = x \cdot 1 = x$
- $\forall x \in G$, $\exists y \in G$ *such that* $y \cdot x = x \cdot y = 1$
- $\forall x, y \in G$, $x \cdot y \in G$
- $\cdot$ *is associative, i.e.* $\forall x, y, z \in G$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.

When working with groups, often the $\cdot$ is ommitted, and we simply write xy instead of $x \cdot y$, where it is understood it is the group operation.
For notation, x^{-1} represents the inverse of $x \in G$. In other words, $x^{-1} \in G$ such that $x^{-1}x = 1$.
It can be shown that x^{-1} is unique in G.

Proof. Let G be a group. Let $x \in G$. Let $y, z \in G$ such that $yx = zx = 1$.

Then

$$
\begin{aligned}
xy &= 1 \\
z(xy) &= z \\
(zx)y &= z \\
y &= z.
\end{aligned}
$$

□

Now we move on to define the special type of Markov Chains we'll be working with.

Definition (Simple Random Walks)**.** *A simple Random Walk is a sequence of random variables $\{X_i\}_{i=0}$, such that for random variables $Z_i, i \in \mathbb{N}$, satisfying $\mathbb{P}(Z_i = 1) = p$ and $\mathbb{P}(Z_i = -1) = 1 - p$ for a $p \in (0, 1)$,*

- $X_0 = 0$,
- $X_n = X_{n-1} + Z_n$, $\forall n \in \mathbb{N}$.

And now, we put them together!

Definition (Groups and random walks)**.** *If $\mathcal{X}$ is a group, then $Q(\cdot)$ is a probability distribution on $\mathcal{X}$, such that $\forall x, y \in \mathcal{X}$, $P(x, y) = Q(x^{-1}y)$.*
Q is called the step distribution.

From here until the end of the chapter, we will assume that $\mathcal{X}$ is a finite Group.

Our first result should help to show that these special types of distributions vastly simplify everything we are studying, reducing the stationary distribution in such cases to arguably the simplest form, uniform.

Proposition 47. *The distribution $\pi(x) = 1/n$, $\forall x \in \mathcal{X}$, where $n = |\mathcal{X}|$, is stationary for every random walk on $\mathcal{X}$ where $\mathcal{X}$ is a finite group.*

Proof. Let P be the transition matrix of the random walk on $\mathcal{X}$. Let $y \in \mathcal{X}$.

Then

$$\begin{aligned}
(\pi P)_y &= \sum_{x\in\mathcal{X}} \pi(x)P(x,y) \\
&= (1/n)\sum_{x\in\mathcal{X}} P(x,y) && \text{(by def. of } \pi) \\
&= (1/n)\sum_{x\in\mathcal{X}} Q(x^{-1}y) && \text{(as } \mathcal{X} \text{ is a group)} \\
&= (1/n)\sum_{z\in\mathcal{X}} Q(z) && \text{(as } \mathcal{X} \text{ is a group)} \\
&= 1/n && \text{(as } Q \text{ is a prob. dis.)} \\
&= \pi(y). && \text{(by def. of } \pi)
\end{aligned}$$

□

An important note here is that all finite abelian groups, groups whose operation is also commutative, can be expressed as the cartesian product of finite groups of modular integers for some base. In other words, given a finite abelian group $\mathcal{X}$, $\exists r \in \mathbb{N}$, $\exists n_0, \ldots, n_r \in \mathbb{Z}$ such that $\mathcal{X} = \mathbb{Z}/(n_0) \times \cdots \times \mathbb{Z}/(n_r)$, where $\mathbb{Z}/(n_i)$ for $i \in \{0, \ldots, r\}$ is the integers mod n_i.
Furthermore, for abelian groups, we use addition and subtraction signs to denote the function of the group. So instead of xy, we would write $x + y$. Similarly for $x^{-1}y$, we would write $y - x$.

We continue with a set of very helpful functions in our discussion of finite groups, characters.

Definition (Characters)**.** *Let $\mathcal{X}$ be a finite abelian group, in the form above. (Here i is the imaginary constant.)*
Then $\forall m = (m_0, \ldots, m_r) \in \mathcal{X}$, let $\chi_m : \mathcal{X} \to \mathbb{C}$ such that $\forall x \in \mathcal{X}$

$$\chi_m(x) := \exp[2\pi i(\frac{m_0 x_0}{n_0} + \cdots + \frac{m_r x_r}{n_r})].$$

Proposition 48 (Facts about characters)**.** *$\forall m, j, x, y \in \mathcal{X}$,*

1. $\chi_m(x + y) = \chi_m(x)\chi_m(y)$.
2. $\chi_m(0) = 1.|\chi_m(x)| = 1.\chi_m(-x) = \overline{\chi_m(x)}$.

3. $\langle \chi_m, \chi_j \rangle = \delta_{mj}$.

4. $\sum_{m\in\mathcal{X}} \chi_m(x) = n\delta_{x0}$.

Proof. The first two items can be easily verified using properties of exponents, Euler's formula, and trigonometric identities.
Let $m, j, x, y \in \mathcal{X}$, such that z_i is the ith coordinate of $z \in \mathcal{X}$ for $z = m, j, x, y$ and $i \in \{0, \dots, r\}$.
Using the first two facts and Proposition 47,

$$\begin{aligned}\langle \chi_m, \chi_j \rangle &= \sum_{x\in\mathcal{X}} \chi_m(x)\overline{\chi_j(x)}\pi(x) \\ &= \sum_{x\in\mathcal{X}} \chi_x(m)\overline{\chi_x(j)}(1/n) \\ &= (1/n)\sum_{x\in\mathcal{X}} \chi_x(m)\chi_x(-j) \\ &= (1/n)\sum_{x\in\mathcal{X}} \chi_x(m-j).\end{aligned}$$

So if $m = j$, then

$$\langle \chi_m, \chi_j \rangle = (1/n)\sum_{x\in\mathcal{X}} \chi_x(m-j) = (1/n)\sum_{x\in\mathcal{X}} \chi_x(0) = (1/n)\sum_{x\in\mathcal{X}} 1 = 1.$$

If $m \neq j$, then

$$\begin{aligned}\langle \chi_m, \chi_j \rangle &= (1/n)\sum_{x\in\mathcal{X}} \chi_x(m-j) \\ &= (1/n)\sum_{x\in\mathcal{X}} \chi_x(k) \\ &\qquad\qquad\qquad\qquad (\text{where } k = m - j \in \mathcal{X}) \\ &= (1/n)[\sum_{x\in\mathcal{X}} \cos(2\pi(\frac{k_0x_0}{n_0} + \cdots + \frac{k_rx_r}{n_r})) + i\sum_{x\in\mathcal{X}} \sin(2\pi(\frac{k_0x_0}{n_0} + \cdots + \frac{k_rx_r}{n_r}))] \\ &= (1/n)[0 + i0] = 0.\end{aligned}$$

So, $\langle \chi_m, \chi_j \rangle = \delta_{mj}$, $\forall m, j \in \mathcal{X}$.

Then for fact 4, notice

$$\sum_{m\in\mathcal{X}} \chi_m(x) = \sum_{m\in\mathcal{X}} \chi_m(x)\chi_m(0)$$

(as $\chi_m(0) = 1$.)

$$= \sum_{m\in\mathcal{X}} \chi_m(x)\overline{\chi_m(0)}$$

(as $0 = -0$, and by item 2. above)

$$= \sum_{m\in\mathcal{X}} \chi_x(m)\overline{\chi_0(m)}$$

(by direct computation)

$$= (n)(1/n)\sum_{m\in\mathcal{X}} \chi_x(m)\overline{\chi_m(0)}$$

$$= (n)\sum_{m\in\mathcal{X}} \chi_x(m)\overline{\chi_m(0)}(1/n)$$

$$= (n)\sum_{m\in\mathcal{X}} \chi_x(m)\overline{\chi_m(0)}\pi(m)$$

(as $\pi(m) = 1/n$, $\forall m \in \mathcal{X}$ is stationary by Proposition 47)

$$= n\delta_{x0}.$$

(by item 3 above)

□

Now on to how characters relate to groups and random walks.

Proposition 49. *$\forall m \in \mathcal{X}$, $\overline{\chi_m}P = \mathbb{E}_Q(\chi_m)\overline{\chi_m} = \lambda_m\overline{\chi_m}$, where $\lambda_m = \mathbb{E}_Q(\chi_m)$, Q is the step distribution, and $\mathbb{E}$ is the expected value function.*

Proof. Let $m \in \mathcal{X}$. Then $\forall y \in \mathcal{X}$,

$$
\begin{aligned}
(\overline{\chi_m}P)_y &= \sum_{x\in\mathcal{X}} \overline{\chi_m}(x)P(x,y) \\
&= \sum_{x\in\mathcal{X}} \chi_m(-x)Q(y-x) && \text{(by def. of } Q) \\
&= \sum_{z\in\mathcal{X}} \chi_m(z-y)Q(z) && \text{(by setting } z=y-x\text{, and as it is a group)} \\
&= \sum_{z\in\mathcal{X}} \chi_m(z)\chi_m(-y)Q(z) && \text{(by Prop. 48)} \\
&= \sum_{z\in\mathcal{X}} \chi_m(z)\overline{\chi_m(y)}Q(z) && \text{(by Prop. 48)} \\
&= \overline{\chi_m(y)}\sum_{z\in\mathcal{X}} \chi_m(z)Q(z) \\
&= \overline{\chi_m(y)}\mathbb{E}_Q(\chi_m). && \text{(by definition of } \mathbb{E})
\end{aligned}
$$

So, as the equality is satisfied for each column, our final equality is satisfied. $\square$

This Proposition shows us that for each $m \in \mathcal{X}$, $\overline{\chi_m}$ is an eigenvector of P for the eigenvalue $\mathbb{E}_Q(\chi_m)$.
This and the fact that $\{\chi_m\}_{m\in\mathcal{X}}$ is an orthonormal basis in $L^2(\pi)$, gives us the following result.

Theorem 50. *For any random walk on any finite abelian group,* $||\mu_k - \pi|| \leq \frac{1}{2}\sqrt{\sum_{j=1}^{n-1} |\lambda_j|^{2k}} \leq (\sqrt{n-1}/2)(\lambda_*)^k$. *In particular,* $\lambda_j = \mathbb{E}_Q(\chi_j)$.

Proof. As $\{\chi_m\}_{m\in\mathcal{X}}$ is an orthonormal basis for $L^2(\pi)$ by Proposition 48, so too must $\{\overline{\chi_m}\}_{m\in\mathcal{X}}$ be an orthonormal basis for $L^2(\pi)$. And by Proposition 49 as each is also an eigenvector, P must be diagonalizable.
Notice further that $\mathbb{E}_Q(\chi_m) < 1$, $\forall m \neq 0$, and $\mathbb{E}_Q(\chi_0) = 1$.
So as $\lambda_j = \mathbb{E}_Q(\chi_j)$ by Proposition 49, by Lemma 7, $\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2\pi(x) = \sum_{j=1}^{n-1} |a_j|^2|\lambda_j|^{2k}$.
Now as we are working with a random walk on a finite group, by Proposition 47, $\pi(x) = 1/n$, $\forall x \in \mathcal{X}$.
Notice that in Theorem 6 and Lemma 7 as $\pi = a_0 v_0$, we get that $a_0 v_0(x) = 1/n$, $\forall x \in \mathcal{X}$.

So, v_0 is the vector such that $v_0(x) = 1/(a_0 n)$, $\forall x \in \mathcal{X}$. So, as the only character satisfying this condition is $\overline{\chi_0}$, we get that $v_0 = \overline{\chi_0}$. So, $a_0 = 1/n$. So, as $\pi(x) = 1/n = a_j$, $\forall j \in \{1, \ldots, n-1\}$,

$$\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2 (1/n) = \sum_{j=1}^{n-1} |(1/n)|^2 |\lambda_j|^{2k}$$

$$\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2 = (1/n) \sum_{j=1}^{n-1} |\lambda_j|^{2k}.$$

Let $u = (1, \ldots, 1) \in \mathbb{R}^n$. Then, as $||\mu_k - \pi|| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|$ as $\mathcal{X}$ is finite,

$$\begin{aligned}
(2||\mu_k - \pi||)^2 &= (\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|)^2 \\
&= |(u \cdot |\mu_k - \pi|)|^2 \\
&\qquad\qquad \text{(where $\cdot$ is the inner product s.t. } u \cdot v = \sum_{x\in\mathcal{X}} u(x)v(x)) \\
&\leq (u \cdot u)(|\mu_k - \pi| \cdot |\mu_k - \pi|) \\
&\qquad\qquad\qquad\qquad \text{(by the Cauchy-Schwarz inequality)} \\
&= (\sum_{x\in\mathcal{X}} 1)(\sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2) \\
&= n \sum_{x\in\mathcal{X}} |\mu_k(x) - \pi(x)|^2 \\
&= (n)(1/n) \sum_{j=1}^{n-1} |\lambda_j|^{2k} \\
&= \sum_{j=1}^{n-1} |\lambda_j|^{2k} \\
&\leq \sum_{j=1}^{n-1} |\lambda_*|^{2k} \\
&= (n-1)|\lambda_*|^{2k}.
\end{aligned}$$

So, taking square root and dividing by two,

$$||\mu_k - \pi|| \leq \frac{1}{2}\sqrt{\sum_{j=1}^{n-1} |\lambda_j|^{2k}} \leq (\sqrt{n-1}/2)(\lambda_*)^k.$$

□

Remark. *Note that in Markov Chains, Eigenvalues and Coupling by Jeffrey Rosenthal, the above theorem is presented as* $||\mu_k - \pi|| \leq \frac{1}{2}\sqrt{\sum_{j=1}^{n-1} |\lambda_j|^{2k}} \leq (\sqrt{n}/2)(\lambda_*)^k$*, though as we found in the above, the slightly stronger inequality, with* $\sqrt{n-1}/2$ *holds.*

So Group space states and Random Walks give us an easy and clear upper bound on the variation distance between the random walk, and the uniform stationary distribution.

11 Coupling and Uniform Minorization Conditions

Stepping away from linear algebra and towards general state spaces, in this chapter we use coupling to attain bounds on the total variation distance of not only finite state space Markov Chains, but of General stat space Markov Chains.

We start with a fact that will be the basis for the idea of coupling.

Proposition 51. *Let* X *and* Y *be two random variables defined on the state space* $\mathcal{X}$*, with probability distributions* $\mathcal{L}(X)$ *and* $\mathcal{L}(Y)$ *respectively. Then* $||\mathcal{L}(X) - \mathcal{L}(Y)|| \leq \mathbb{P}(X \neq Y)$*.*

Proof.

$$\begin{aligned}
||\mathcal{L}(X) - \mathcal{L}(Y)|| :&= \sup_{A\subseteq\mathcal{X}}|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\
&= \sup_{A\subseteq\mathcal{X}}|\mathbb{P}(X \in A, X = Y) + \mathbb{P}(X \in A, X \neq Y) \\
&\qquad - \mathbb{P}(Y \in A, X = Y) - \mathbb{P}(Y \in A, X \neq Y)| \\
&= \sup_{A\subseteq\mathcal{X}}|\mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y)| \\
&\qquad \text{(as if } X = Y\text{, then } X, Y \in A) \\
&\leq \mathbb{P}(X \neq Y).
\end{aligned}$$

□

So. given a Markov Chain $\{X_i\}_{i=0}$ with transition kernel $P(\cdot,\cdot)$, state space $\mathcal{X}$, and initial distribution μ_0, if we find another Markov Chain $\{(X_i, Y_i)\}_{i=0}$ on the state space $\mathcal{X} \times \mathcal{X}$ where

- $\forall i \in \mathbb{N}$, X_i, Y_i follow the transition kernel $P(\cdot,\cdot)$,
- $X_0 \sim \mu_0$,
- $Y_0 \sim \pi$ where π is a stationary distribution,
- there exists a random variable T such that $\forall k \in \mathbb{N}$ such that $k \geq T$, $X_k = Y_k$,

then $||\mu_k - \pi|| \leq \mathbb{P}(T > k)$. T is called the coupling time, and any Markov Chain with these properties is called a coupling.
We shall see how in the following:

Theorem 52. *Let there be Markov Chains as in the above statement. Then* $||\mu_k - \pi|| \leq \mathbb{P}(T > k)$.

Proof. As $\{X_i\}_{i=0}$ follows the transition kernel $P(\cdot,\cdot)$, $\forall k \in \mathbb{N}$, $\mathcal{L}(X_k) = \mu_k$ by definition of μ_k.
As $\{Y_i\}_{i=0}$ follows the same transition kernel, and starts at the stationary distribution π, $\forall k \in \mathbb{N}$, $\mathcal{L}(Y_k) = \pi$ as π is stationary.
So, $||\mu_k - \pi|| = ||\mathcal{L}(X_k) - \mathcal{L}(Y_k)|| \leq \mathbb{P}(X_k \neq Y_k) \leq \mathbb{P}(T > k)$, by Proposition 51 and as $\forall k \geq T$, $X_k = Y_k$. □

We can use coupling to find bounds on the variation distance with the stopping time, T, however finding such a coupling is not always easy.

We now define and use a condition which we can use to find such a coupling.

Definition (Uniform Minorization Conditions)**.** *A minorization condition for a Markov Chain is the existence of a* $\beta \in (0,1]$ *and a probability distribution on* $\mathcal{X}$, ζ, *such that* $\exists k_0 \in \mathbb{N}$ *such that for every* $x \in \mathcal{X}$ *and for every measurable subset* $A \subseteq \mathcal{X}$,

$$P^{k_0}(x, A) \geq \beta\zeta(A).$$

This Uniform Minorization Condition gives us a coupling to produce the following.

Theorem 53. *Given a Markov Chain on a state space $\mathcal{X}$ with transition kernel $P(\cdot,\cdot)$, if there exists $\beta \in (0,1]$ and a probability distribution ζ on $\mathcal{X}$ such that $P(x,A) \geq \beta\zeta(A)$ for all $x \in \mathcal{X}$ and all measurable subsets $A \subseteq \mathcal{X}$, then given any initial distribtution μ_0 and any stationary distribution π, $||\mu_k - \pi|| \leq (1-\beta)^k$.*

Proof. Let $\mathcal{X}$ be a state space, $P(\cdot,\cdot)$ be a transition kernel on $\mathcal{X}$, μ_0 a probability distribution on $\mathcal{X}$, and π be a stationary distribution on $\mathcal{X}$.
Let $\beta \in (0,1]$ and ζ be a probability distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}$ and for all measurable subsets $A \subseteq \mathcal{X}$, $P(x,A) \geq \beta\zeta(A)$.
Let $\{(X_i, Y_i)\}_{i=0}$ be a Markov Chain as follows.
Let $X_0 \perp\!\!\!\perp Y_0$ (X_0 be independent from Y_0) such that $X_0 \sim \mu_0$ and $Y_0 \sim \pi$.
Then let $\{W_i\}_{i=1}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(W_i = 1) = \beta$, and $\mathbb{P}(W_i = 0) = 1-\beta$, $\forall i$.
Then $\forall k \in \mathbb{N}$, if $W_k = 1$, choose $z \in \mathcal{X}$ randomly by ζ, and set $z = X_k = Y_k$.
So, $\mathbb{P}(X_k \in A | W_k = 1) = \zeta(A)$, and similarly for Y_k.
If $W_k = 0$, then pick X_k and Y_k such that $X_k \perp\!\!\!\perp Y_k$, and

$$\mathbb{P}(X_k \in A) = \frac{P(X_{k-1},A) - \beta\zeta(A)}{1-\beta}, \mathbb{P}(Y_k \in A) = \frac{P(Y_{k-1},A) - \beta\zeta(A)}{1-\beta}.$$

Notice we can only do this because we know $P(x,A) \geq \beta\zeta(A)$, $\forall x \in \mathcal{X}$ and all measureable subsets $A \subseteq \mathcal{X}$, as this ensures $\mathbb{P}(X_k \in A) = \frac{P(X_{k-1},A)-\beta\zeta(A)}{1-\beta} \geq 0$, $\forall k \in \mathbb{N}$, and similarly for $\{Y_i\}_{i=0}$.
So for X_0 as well as Y_0,

$$\mathbb{P}(X_k \in A | W_k = 0) = \frac{P(X_{k-1},A) - \beta\zeta(A)}{1-\beta}.$$

Then, $\forall k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(X_k \in A | X_{k-1}) &= \mathbb{P}(W_k = 1)\mathbb{P}(X_k \in A | W_k = 1) + \mathbb{P}(W_k = 0)\mathbb{P}(X_k \in A | W_k = 0) \\
&= \mathbb{P}(W_k = 1)\zeta(A) + \mathbb{P}(W_k = 0)\frac{P(X_{k-1},A) - \beta\zeta(A)}{1-\beta} \\
&= \beta\zeta(A) + (1-\beta)\frac{P(X_{k-1},A) - \beta\zeta(A)}{1-\beta} \\
&= P(X_{k-1},A).
\end{aligned}$$

Similary for Y_k, $\mathbb{P}(Y_k \in A|Y_{k-1}) = P(Y_{k-1}, A)$.
Let T be the first k such that $W_k = 1$.
Then define a new Markov Chain $\{(X_i, Z_i)\}_{i=0}$ on $\mathcal{X} \times \mathcal{X}$, such that

$$Z_k = \begin{cases} Y_k, & \text{if } k < T \\ X_k, & \text{if } k \geq T \end{cases}.$$

Then

$$\begin{aligned}
\mathbb{P}(Z_k \in A|Z_{k-1}) &= \mathbb{P}(T > k)\mathbb{P}(X_k \in A|X_{k-1}) + \mathbb{P}(T \leq k)\mathbb{P}(Y_k \in A|Y_{k-1}) \\
&= \mathbb{P}(T > k)P(X_{k-1}, A) + \mathbb{P}(T \leq k)P(Y_{k-1}, A) \\
&= \mathbb{P}(T > k)P(Z_{k-1}, A) + \mathbb{P}(T \leq k)P(Z_{k-1}, A) \\
&\text{(as when } T \leq k,\ Z_{k-1} = Y_{k-1}, \text{ and when } T > k,\ Z_{k-1} = X_{k-1}) \\
&= P(Z_{k-1}, A).
\end{aligned}$$

So, as $X_0 \sim \mu_0$, $Z_0 \sim \pi$, $\mathbb{P}(X_k \in A|X_{k-1}) = P(X_{k-1}, A)$, $\mathbb{P}(Z_k \in A|Z_{k-1}) = P(Z_{k-1}, A)$, and $\forall k \geq T$, $X_k = Z_k$, the Markov Chain $\{(X_k, Z_k)\}_{i=0}$ is a coupling as defined earlier.
So by Theorem 52,

$$\begin{aligned}
||\mu_k - \pi|| &\leq \mathbb{P}(T > k) \\
&= \mathbb{P}(W_1, \ldots, W_k = 0) \\
&= \mathbb{P}(W_1 = 0) \cdots \mathbb{P}(W_k = 0) \\
&= (\mathbb{P}(W_1 = 0))^k = (1 - \beta)^k.
\end{aligned}$$

□

This theorem gives us a bound on the variation distance of any Markov Chain that satisfies a uniform minorization condition, restricted to the first step. Later the result will be generalized to any uniform minorization condition without too much trouble.

But for now, let us go back to the finite case. The above theorem allows us to require only a column of the matrix P to be greater than zero in order to find a bound.

Proposition 54. *Given a Markov Chain $\{X_i\}_{i=0}$ on a finite state space $\mathcal{X}$ with transition kernel $P(\cdot, \cdot)$ and transition matrix P, if a column of P has* *all positive entries, then the Markov Chain satisfies $P(x,A) \geq \beta\zeta(A)$ for some $\beta \in (0,1]$ and probability distribution on $\mathcal{X}$, ζ, for every $x \in \mathcal{X}$ and for every measureable subset $A \subseteq \mathcal{X}$.*

Proof. Let $i \in \{0, \dots, n-1\}$ be the column such that all entries are positive, and let $y = x_i$.
Then let $\zeta : \mathcal{X} \to [0,1]$ such that $\forall x \in \mathcal{X}$, $\zeta(x) = \delta_{yx}$.
Then obviously ζ is a probability distribution on $\mathcal{X}$.
Furthermore, as $P(x,y) > 0$, $\forall x \in \mathcal{X}$, let $\beta \in (0, \min_{x \in \mathcal{X}} P(x,y)]$.
Let A be any subset of $\mathcal{X}$.
Then if $y \in A$, then $\beta\zeta(A) = \beta \sum_{x \in \mathcal{X}} \zeta(x) = \beta$.
So, $\forall x \in \mathcal{X}$, $P(x,A) = \sum_{z \in A} P(x,z) \geq P(x,y) \geq \beta = \beta\zeta(A)$, by construction of β.
If $y \notin A$, then $\beta\zeta(A) = \beta \sum_{z \in A} \zeta(z) = 0$.
So, $\forall x \in \mathcal{X}$, $P(x,A) \geq 0 = \beta\zeta(A)$. $\square$

So, given any Markov Chain on a finite state space $\mathcal{X}$, if P has a column of positive entries, then the total variation distance is bounded by $(1-\beta)^k$, for some $\beta \in (0,1]$.

Now we show why $\beta = \int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x,dy)$ is the largest beta we can use.

Proposition 55. *The largest such β that can be used is $\beta = \int_{\mathcal{X}} inf_{x \in \mathcal{X}} P(x,dy)$.*

Proof. First we show such a β satisfies the condition.
Let $\beta = \int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x,dy)$. Let ζ be a function such that for every measurable subset $A \subseteq \mathcal{X}$,

$$\zeta(A) = \frac{\int_A \inf_{x \in \mathcal{X}} P(x,dy)}{\int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x,dy)}.$$

Notice that $\zeta(\mathcal{X}) = 1$, $\zeta(\emptyset) = 0$, and for every measurable subset $A \subseteq \mathcal{X}$ $\zeta(A) \geq 0$.
So by the linearity of integrals, ζ is a probability distribution on $\mathcal{X}$.
Then $\forall x \in \mathcal{X}$ and for every measurable subset $A \subseteq \mathcal{X}$,

$$\begin{aligned}
P(x,dy) &\geq \inf_{j \in \mathcal{X}} P(j,dy) && \text{(by def. of inf)}\\
\int_A P(x,dy) &\geq [\int_A \inf_{j \in \mathcal{X}} P(j,dy)] \frac{\int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x,dy)}{\int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x,dy)} && \text{(by monotonicity of integration)}\\
P(x,A) &\geq \beta\zeta(A). && \text{(by definitions)}
\end{aligned}$$

So $\beta = \int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x, dy)$ satisfies the inequality.
Now let $\beta' \in (0, 1]$ and let ζ' be another probability distribution on $\mathcal{X}$ such that $\beta'\zeta'(A) \leq P(x, A)$ for every $x \in \mathcal{X}$ and for every measureable subset $A \subseteq \mathcal{X}$. Then notice

$$\beta' = \int_{\mathcal{X}} \inf_{x \in \mathcal{X}} \beta'\zeta'(dy)$$

$$\text{(as } \beta' \text{ and } \zeta' \text{ don't depend on } x\text{, and } \int_{\mathcal{X}} \zeta'(dy) = 1 \text{ as } \zeta' \text{ is a prob. dis.)}$$

$$\leq \int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x, dy).$$

$$\text{(by assumption)}$$

So every β that satisfies this uniform minorization condition is less than or equal to $\int_{\mathcal{X}} \inf_{x \in \mathcal{X}} P(x, dy)$, making it the maximum. $\square$

We now prove that the total variation distance is non-increasing, in efforts to generalize Theorem 53. Note here we only present the finite case, but the following Proposition holds for general state space chains as well.

Proposition 56. *The distance to stationarity is weakly decreasing. I.e. for any Markoc chain P and any $k \geq 0$, if π is a stationary distribution of P, then $||\mu_{k+1} - \pi|| \leq ||\mu_k - \pi||$. (Finite case)*

Proof. Let $k \geq 0$.

As $\mathcal{X}$ is finite,

$$
\begin{aligned}
||\mu_{k+1} - \pi||_{\text{var}} &= \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu_{k+1}(x) - \pi(x)| \\
&\qquad \text{(by Proposition 1)} \\
&= \frac{1}{2}\sum_{x\in\mathcal{X}} |\sum_{y\in\mathcal{X}} P(y,x)\mu_k(y) - \sum_{y\in\mathcal{X}} P(y,x)\pi(y)| \\
&\qquad \text{(as } \mathcal{X} \text{ is discrete and } \pi \text{ is a stationary distribution)} \\
&= \frac{1}{2}\sum_{x\in\mathcal{X}} |\sum_{y\in\mathcal{X}} P(y,x)(\mu_k(y) - \pi(y))| \\
&\leq \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} P(y,x)|\mu_k(y) - \pi(y)| \\
&\qquad \text{(by the triangle inequality)} \\
&= \frac{1}{2}\sum_{y\in\mathcal{X}}\sum_{x\in\mathcal{X}} P(y,x)|\mu_k(y) - \pi(y)| \\
&= \frac{1}{2}|\sum_{y\in\mathcal{X}} |\mu_k(y) - \pi(y)|(\sum_{x\in\mathcal{X}} P(y,x)) \\
&= \frac{1}{2}|\sum_{y\in\mathcal{X}} |\mu_k(y) - \pi(y)| \\
&\qquad \text{(as } \sum_{x\in\mathcal{X}} P(y,x)) = 1) \\
&= ||\mu_k - \pi||_{\text{var}}.
\end{aligned}
$$

□

Theorem 57 (Generalization of convergence given uniform minorization conditions (Ex 6))**.** *Given a Markov Chain with state space $\mathcal{X}$, transition kernel $P(\cdot,\cdot)$, any initial distribution μ_0 and any stationary distribution π, if there exists $\beta > 0$ and a probability distribution ζ on $\mathcal{X}$ such that for every $x \in \mathcal{X}$ and for every measurable subset $A \subseteq \mathcal{X}$, $P^{k_0}(x, A) \geq \beta\zeta(A)$ for some $k_0 \in \mathbb{N}$, then $||\mu_k - \pi|| \leq (1-\beta)^{\lfloor k/k_0 \rfloor}$, where $\lfloor k/k_0 \rfloor$ is the largest integer not surpassing k/k_0.*

Proof. Let $k_0 \in \mathbb{N}^+$, μ_0 be an initial disrtibution, and π be a stationary distribution.

Let $X_0 \sim \mu_0$ and $Y_0 \sim \pi$ such that $X_0 \perp\!\!\!\perp Y_0$.
Let $\{W_i\}_{i=1}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(W_k = 1) = \beta$ and $\mathbb{P}(W_k = 0) = 1 - \beta \ \forall k \in \mathbb{N}^+$.
Then $\forall k \in \mathbb{N}$ such that $k_0 \nmid k$, choose X_k such that $\mathbb{P}(X_k \in A|X_{k-1}) = P(X_{k-1}, A)$ for every measurable subset $A \subseteq \mathcal{X}$. Choose Y_k similarly.
Then $\forall k \in \mathbb{N}^+$ such that $k_0|k$, choose X_k and Y_k such that
if $W_{k/k_0} = 1$, choose a $z \in \mathcal{X}$ according to ζ, so $X_k = Y_k = z$, so $\mathbb{P}(X_k \in A|W_{k/k_0} = 1) = \zeta(A)$, and similarly for Y_k,
and if $W_{k/k_0} = 0$, then choose X_k and Y_k such that for any measurable subset $A \subseteq \mathcal{X}$,

$$\mathbb{P}(X_k \in A) = \frac{P^{k_0}(X_{k-k_0}, A) - \beta\zeta(A)}{1-\beta}, \mathbb{P}(Y_k \in A) = \frac{P^{k_0}(Y_{k-k_0}, A) - \beta\zeta(A)}{1-\beta},$$

so $\mathbb{P}(X_k \in A|W_{k/k_0} = 0) = \frac{P^{k_0}(X_{k-k_0},A)-\beta\zeta(A)}{1-\beta}$, and similarly for Y_k.
Then $\forall k \in \mathbb{N}^+$ such that $k_0|k$, $\mathbb{P}(X_k \in A|X_{k-k_0}) = \mathbb{P}(W_{k/k_0} = 1)\mathbb{P}(X_k \in A|W_{k/k_0} = 1) + \mathbb{P}(W_{k/k_0} = 0)\mathbb{P}(X_k \in A|W_{k/k_0} = 0) = \beta\zeta(A) + (1 - \beta)\frac{P^{k_0}(X_{k-k_0},A)-\beta\zeta(A)}{1-\beta} = P^{k_0}(X_{k-k_0}, A)$.
So $\mathbb{P}(X_k \in A|X_{k-k_0}) = P^{k_0}(X_{k-k_0}, A)$, and similarly for Y_k.
As after X_{k-k_0}, X_{j-k_0} gets updated by $P(\cdot, \cdot)$ for $k \le j < k$, and $\mathbb{P}(X_k \in A|X_{k-k_0}) = P^{k_0}(X_{k-k_0}, A)$ for every measurable set $A \subseteq \mathcal{X}$, it follows that $\mathbb{P}(X_k \in A|X_{k-1}) = P(X_k, A)$, and similarly for Y_k.
So, let T be the first $k \in \mathbb{N}$ such that $W_k = 1$.
Then let $\{(X_k, Z_k)\}_{i=0}$ be the Markov Chain on $\mathcal{X} \times \mathcal{X}$ such that $\forall k \in \mathbb{N}$,
$Z_k = \begin{cases} Y_k, & \text{if } k < T \\ X_k, & \text{if } k \ge T \end{cases}$.
Then

$$\begin{aligned}\mathbb{P}(Z_k \in A|Z_{k-1}) &= \mathbb{P}(T > k)\mathbb{P}(X_k \in A|X_{k-1}) + \mathbb{P}(T \le k)\mathbb{P}(Y_k \in A|Y_{k-1}) \\ &= \mathbb{P}(T > k)P(X_{k-1}, A) + \mathbb{P}(T \le k)P(Y_{k-1}, A) \\ &= \mathbb{P}(T > k)P(Z_{k-1}, A) + \mathbb{P}(T \le k)P(Z_{k-1}, A) \\ &\text{(as when } T \le k,\ Z_{k-1} = Y_{k-1}, \text{ and when } T > k,\ Z_{k-1} = X_{k-1}) \\ &= P(Z_{k-1}, A).\end{aligned}$$

So, as $X_0 \sim \mu_0$, $Z_0 \sim \pi$, $\mathbb{P}(X_k \in A|X_{k-1}) = P(X_{k-1}, A)$, $\mathbb{P}(Z_k \in A|Z_{k-1}) = P(Z_{k-1}, A)$, and $\forall k \ge T$, $X_k = Z_k$, the Markov Chain $\{(X_k, Z_k)\}_{i=0}$ is a coupling as defined earlier.

So by Theorem 52, $||\mu_k - \pi|| \le \mathbb{P}(T > k) = \mathbb{P}(W_1, \ldots, W_{k/k_0} = 0) = \mathbb{P}(W_1 = 0) \ldots \mathbb{P}(W_{k/k_0} = 0) = (\mathbb{P}(W_1 = 0))^{k/k_0} = (1-\beta)^{k/k_0}$.
For $k \in \mathbb{N}$ such that $k_0 \nmid k$, by Proposition 56, as the largest integer multiple of k_0 not exceeding k, $\lfloor k/k_0 \rfloor$, satisfies $||\mu_{\lfloor k/k_0 \rfloor} - \pi|| \le (1-\beta)^{\lfloor k/k_0 \rfloor}$, as $k \ge \lfloor k/k_0 \rfloor$, and X_j such that $\lfloor k/k_0 \rfloor \le j \le k$ get updated by $P(\cdot,\cdot)$, $||\mu_k - \pi|| \le ||\mu_{\lfloor k/k_0 \rfloor} - \pi|| \le (1-\beta)^{\lfloor k/k_0 \rfloor}$. ◻

12 Bounds on the Total Variation Distance of Finite Product Measures

For this section, let $\mathcal{X}_1$ and $\mathcal{X}_2$ be finite state spaces. Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$.
Let μ_1 and v_1 be probability distributions on $\mathcal{X}_1$, and let μ_2 and v_2 be probability distributions on $\mathcal{X}_2$.
Let $\mu_1 \times \mu_2$ be the probability distribution on $\mathcal{X}$ such that $(\mu_1 \times \mu_2)(x,y) := \mu_1(x)\mu_2(y)$, $\forall (x,y) \in \mathcal{X}$, and similarly let $v_1 \times v_2$ be the probability distribution on $\mathcal{X}$ such that $(v_1 \times v_2)(x,y) := v_1(x)v_2(y)$, $\forall (x,y) \in \mathcal{X}$

Proposition 58 (Upper Bound)**.**

$$||\mu_1 - v_1||_{var} + ||\mu_2 - v_2||_{var} \le 2||\mu_1 \times \mu_2 - v_1 \times v_2||_{var}.$$

Proof. As $\mathcal{X}_1$ and $\mathcal{X}_2$ are finite sample spaces,

$$||\mu_1 - v_1||_{\mathrm{var}} + ||\mu_2 - v_2||_{\mathrm{var}} = \frac{1}{2}(\sum_{x \in \mathcal{X}_1} |\mu_1(x) - v_1(x)| + \sum_{y \in \mathcal{X}_2} |\mu_2(y) - v_2(y)|).$$

So,

$$\frac{1}{2}(\sum_{x\in\mathcal{X}_1}|\mu_1(x)-v_1(x)|+\sum_{y\in\mathcal{X}_2}|\mu_2(y)-v_2(y)|)$$
$$=\frac{1}{2}(\sum_{x\in\mathcal{X}_1}|\sum_{y\in\mathcal{X}_2}\mu_1(x)\mu_2(y)-\sum_{y\in\mathcal{X}_2}v_1(x)v_2(y)|+\sum_{y\in\mathcal{X}_2}|\sum_{x\in\mathcal{X}_1}\mu_1(x)\mu_2(y)-\sum_{x\in\mathcal{X}_1}v_1(x)v_2(y)|)$$
$$\text{(as they are probability distributions, so } 1=\sum_{s\in S}\mu(s))$$
$$=\frac{1}{2}(\sum_{x\in\mathcal{X}_1}|\sum_{y\in\mathcal{X}_2}(\mu_1(x)\mu_2(y)-v_1(x)v_2(y))|+\sum_{y\in\mathcal{X}_2}|\sum_{x\in\mathcal{X}_1}(\mu_1(x)\mu_2(y)-v_1(x)v_2(y))|)$$
$$\leq\frac{1}{2}(\sum_{x\in\mathcal{X}_1}\sum_{y\in\mathcal{X}_2}|\mu_1(x)\mu_2(y)-v_1(x)v_2(y)|+\sum_{y\in\mathcal{X}_2}\sum_{x\in\mathcal{X}_1}|\mu_1(x)\mu_2(y)-v_1(x)v_2(y)|)$$
$$\text{(by the triangle inequality)}$$
$$=\frac{1}{2}(\sum_{(x,y)\in\mathcal{X}}|\mu_1(x)\mu_2(y)-v_1(x)v_2(y)|+\sum_{(x,y)\in\mathcal{X}}|\mu_1(x)\mu_2(y)-v_1(x)v_2(y)|)$$
$$\text{(grouping sums)}$$
$$=(2)(\frac{1}{2})\sum_{(x,y)\in\mathcal{X}}|\mu_1(x)\mu_2(y)-v_1(x)v_2(y)|$$
$$=2||\mu_1\times\mu_2-v_1\times v_2||_{\text{var}}.$$

□

Proposition 59 (Lower Bound)**.**

$$||\mu_1\times\mu_2-v_1\times v_2||_{var}\leq||\mu_1-v_1||_{var}+||\mu_2-v_2||_{var}.$$

Proof. As $\mathcal{X}_1$ and $\mathcal{X}_2$ are finite state spaces, $\mathcal{X}_1\times\mathcal{X}_2=\mathcal{X}$ is a finite state space.
So, as $\mathcal{X}$ is a finite state space, $||\mu_1\times\mu_2-v_1\times v_2||_{\text{var}}=\frac{1}{2}\sum_{(x,y)\in\mathcal{X}}|\mu_1(x)\mu_2(y)-v_1(x)v_2(y)|$.

So,

$$
\begin{aligned}
&\frac{1}{2}\sum_{(x,y)\in\mathcal{X}} |\mu_1(x)\mu_2(y) - v_1(x)v_2(y)| \\
&= \frac{1}{2}\sum_{(x,y)\in\mathcal{X}} |\mu_1(x)\mu_2(y) - \mu_1(x)v_2(y) + \mu_1(x)v_2(y) - v_1(x)v_2(y)| \\
&= \frac{1}{2}\sum_{(x,y)\in\mathcal{X}} |\mu_1(x)(\mu_2(y) - v_2(y)) + v_2(y)(\mu_1(x) - v_1(x))| \\
&\leq \frac{1}{2}\sum_{(x,y)\in\mathcal{X}} (|\mu_1(x)(\mu_2(y) - v_2(y))| + |v_2(y)(\mu_1(x) - v_1(x))| \\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(by the triangle inequality)} \\
&= \frac{1}{2}(\sum_{(x,y)\in\mathcal{X}} |\mu_1(x)(\mu_2(y) - v_2(y))| + \sum_{(x,y)\in\mathcal{X}} |v_2(y)(\mu_1(x) - v_1(x))|) \\
&= \frac{1}{2}(\sum_{(x,y)\in\mathcal{X}} \mu_1(x)|\mu_2(y) - v_2(y)| + \sum_{(x,y)\in\mathcal{X}} v_2(y)|\mu_1(x) - v_1(x)|) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(as } \mu \geq 0 \text{ as a prob. dis.)} \\
&= \frac{1}{2}(\sum_{y\in\mathcal{X}_2}\sum_{x\in\mathcal{X}_1} \mu_1(x)|\mu_2(y) - v_2(y)| + \sum_{x\in\mathcal{X}_1}\sum_{y\in\mathcal{X}_2} v_2(y)|\mu_1(x) - v_1(x)|) \\
&= \frac{1}{2}(\sum_{y\in\mathcal{X}_2} |\mu_2(y) - v_2(y)| + \sum_{x\in\mathcal{X}_1} |\mu_1(x) - v_1(x)|) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(as } \sum_{s\in S}\mu = 1 \text{ as prob. dis.)} \\
&= ||\mu_1 - v_1||_{\text{var}} + ||\mu_2 - v_2||_{\text{var}}.
\end{aligned}
$$

□

Combining these bounds, we get the following.

Theorem 60.

$||\mu_1 \times \mu_2 - v_1 \times v_2||_{var} \leq ||\mu_1 - v_1||_{var} + ||\mu_2 - v_2||_{var} \leq 2||\mu_1 \times \mu_2 - v_1 \times v_2||_{var}.$

Examples.

1. *Trivial Example:* $\mathcal{X}_1 = \{x_{11}, \ldots, x_{1n}\}$, $\mathcal{X}_2 = \{x_{21}, \ldots, x_{2m}\}$, μ_1 *and* v_1 *are probability distributions on* $\mathcal{X}_1$ *such that* $\mu_1(x) = v_1(x)$, $\forall x \in \mathcal{X}_1$, *and* μ_2 *and* v_2 *are probability disrtibutions on* $\mathcal{X}_2$ *such that* $\mu_2(x) = v_2(x)$, $\forall x \in \mathcal{X}_2$. *Then*

   $$||\mu_1 \times \mu_2 - v_1 \times v_2|| = ||\mu_1 - v_1|| + ||\mu_2 - v_2|| = 2||\mu_1 \times \mu_2 - v_1 \times v_2|| = 0.$$

2. $\mathcal{X}_1 = \mathcal{X}_2 = \{a, b\}$, $\mu_1 = \mu_2 = (1, 0)$, $v_1 = v_2 = (0, 1)$. *Then*

   $$1 = ||\mu_1 \times \mu_2 - v_1 \times v_2|| = ||\mu_1 - v_1|| + ||\mu_2 - v_2|| < 2||\mu_1 \times \mu_2 - v_1 \times v_2|| = 2.$$

3. $\mathcal{X}_1 = \mathcal{X}_2 = \{a, b\}$, $\mu_1 = \mu_2 = (3/4, 1/4)$, $v_1 = v_2 = (1/4, 3/4)$. *Then*

   $$1/2 = ||\mu_1 \times \mu_2 - v_1 \times v_2|| < ||\mu_1 - v_1|| + ||\mu_2 - v_2|| = 2||\mu_1 \times \mu_2 - v_1 \times v_2|| = 1.$$

4. $\mathcal{X}_1 = \{x_{11}, x_{12}\}$, $\mathcal{X}_2 = \{x_{21}, x_{22}, x_{23}\}$, $\mu_1 = (3/4, 1/4)$, $v_1 = (1, 0)$, $\mu_2 = (1/3, 1/3, 1/3)$, $v_2 = (1/3, 0, 2/3)$. *Then*

   $$\begin{aligned} ||\mu_1 \times \mu_2 - v_1 \times v_2|| &= 11/24 < \\ ||\mu_1 - v_1|| + ||\mu_2 - v_2|| = 1/4 + 1/3 &= 7/12 < \\ 2||\mu_1 \times \mu_2 - v_1 \times v_2|| &= 11/12. \end{aligned}$$

13 Appendix

Lemma 61. *If a finite state space Markov Chain is indecomposable and aperiodic, then* $\forall x, y \in \mathcal{X}$, $\exists k_x \in \mathbb{N}$ *such that* $\forall k \geq k_x$, $P^k(x, y) > 0$.

Proof. Let $x, y \in \mathcal{X}$. Let $S_x = \{k \in \mathbb{N} : P^k(x, x) > 0\}$.
As the Markov Chain is aperiodic, by definition $\gcd(S_x) = 1$.
As $\gcd(S_x) = 1$, by the Euclidean Algorithm, $\exists m \in \mathbb{N}$, $a_0, \ldots, a_m \in \mathbb{Z}$, and $s_0, \ldots, s_m \in S_x$ such that $1 = a_0 s_0 + \cdots + a_m s_m$.
Let $M = |a_0| s_0 + \cdots + |a_m| s_m$. Let $p \in S_x$, and let $q \in \mathbb{N}$ such that $P^q(x, y) > 0$ (guaranteed by irreducibility).
Then let $k_x = pM + q$. Then let $k \in \mathbb{N}$ such that $k \geq k_x$.
Then let $c \in \mathbb{N}$ such that c is the largest integer such that $p|c$ and $c \leq k - q$.
Then $k = pc + q + r$, where $0 \leq r < p$ (think of as a remainder), and $c \geq M$

(as $k \geq k_x$, and if $k = k_x$, then $c = M$).
So,

$$\begin{aligned} k &= pc + q + r \\ &= pc - pM + pM + r(\sum_{i=0}^{m} a_i s_i) + q && (\text{as } \sum_{i=0}^{m} a_i s_i = 1) \\ &= p(c - M) + p(\sum_{i=0}^{m} |a_i| s_i) + r(\sum_{i=0}^{m} a_i s_i) + q && (\text{by def. of M}) \\ &= p(c - M) + \sum_{i=0}^{m} [(p|a_i| + r a_i) s_i] + q. \end{aligned}$$

Using the fact for any $i > 1$ and any $i > j \geq 1$, $P^i(x, y) = \sum_{x \in \mathcal{X}} P^j(x, x) P^{i-j}(x, y)$,

$$P^k(x, y) \geq (P^p(x, x))^{(c-M)} [\prod_{i=0}^{m} (P^{s_i}(x, x))^{(p|a_i| + r a_i)}] (P^q(x, y)) > 0.$$

As k is arbitrary except that $k \geq k_x$, the result is proven. $\square$

Lemma 62. *If P is irreducible and reversible wrt π, then $\lambda_i < 1$ for every $i \in \{1, \ldots, n-1\}$.*

Proof. By Proposition 4, $\lambda_i \leq 1$ for every i.
So assume for a contradiction that $\lambda_1 = 1$. As P is reversible, by Lemma 11, P is diagonalizable.
As P is diagonalizable, and $\lambda_0 = \lambda_1 = 1$, the eigenspace of 1, $E(1)$, has at dimension at least 2.
So, as π is a stationary distribution, $\pi P = \pi$. (This is because for $0 \leq j \leq n-1$, $\pi_j = \sum_{i=0}^{n-1} \pi(x_i) P(x_i, x_j) = (\pi P)_j$).
So, $\exists v \in \mathbb{R}^{\mathcal{X}}$ such that $\{\pi, v\}$ is linearly independent and span$(\{\pi, v\}) \subseteq E(1)$.
As $\pi(y) > 0$, $\forall y \in \mathcal{X}$ by Proposition 9, let $M \in \mathbb{R}$ such that $M\pi(x) > |v(x)|$, $\forall x \in \mathcal{X}$.
Then let $w \in \mathbb{R}^{\mathcal{X}}$ such that $w = M\pi + v$, so $w(x) > 0$, $\forall x \in \mathcal{X}$. Now let $\mu : 2^{\mathcal{X}} \to \mathbb{R}$ such that

$$\mu(A) = \frac{\sum_{x \in A} w(x)}{\sum_{x \in \mathcal{X}} w(x)}, \qquad \forall A \in 2^{\mathcal{X}}.$$

Notice that μ is now a probability distribution on $\mathcal{X}$.
Let $C = 1/[\sum_{x \in \mathcal{X}} w(x)]$. Then notice

$$\mu P = (Cw)P = C(M\pi + v)P = C(M(\pi P) + vP) = C(M\pi + v) = Cw = \mu,$$

so μ is also a stationary distribution.
And as $\mu = Cw = C(M\pi + v)$, and v and π are linearly independent, μ and π are linearly independent, and hence not equal.
But, as P is an irreducible Markov Chain, P can have at max one stationary distribution, and thus we have arrived at a contradiction. □

If instead of wanting to believe that an irreducible Markov Chain has at most one stationary distribution, we can instead introduce the Perron-Frobenius Theorem for non-negative matrices, in order to show Lemma 62.

Theorem 63 (Perron-Frobenius)**.** *If $X \in M^{n \times n}(\mathbb{R})$ such that each entry of X is non-negative, and X is irreducible, then the maximum eigenvalue of X, call it λ_0, satisfies $\lambda_0 \geq |\lambda|$ for all other eigenvalues λ of X, and λ_0 has algebraic multiplicity one.*

Easier proof of Lemma 62.

Proof. As P is irreducible, nonnegative, square, and only has real entries, by Perron-Frobenius, the max eigenvalue, which by Proposition 3 and 4 is $\lambda_0 = 1$, has algebraic multiplicity one. As P is reversible, it is diagonalizable, and thus λ_0 also has geometric multiplicity equal to one.
So for every other eigenvalue λ of P, as $|\lambda| \leq \lambda_0$ and $\lambda \neq \lambda_0$ as it has multiplicity one, $\lambda < \lambda_0 = 1$. □

14 References

- *A First Look at Stochastic Processes*, Jeffrey S. Rosenthal, 2019
- *Efficiency of Reversible MCMC Methods: Elementary Derivations and Applications to Composite Methods*, Radford M. Neal, Jeffrey S. Rosenthal, 2023
- *Markov Chains, Eigenvalues, and Coupling*, Jeffrey S. Rosenthal, 1993
- *Matrix Analysis, Second Edition*, Roger A. Horn, Charles R. Johnson, 1985 or 2013?

- Chapter 11 *Stochastic Processes* in *Probability and Statistics*, Michael J. Evans and Jeffrey S. Rosenthal, 2009