Efficiency Dominance Generalization

Geoffrey Salmon, Supervised by Jeffrey Rosenthal

May 4, 2024

1 Setup

Setup: $\mathbf{X}$ is a nonempty set and $\mathcal{F}$ is a σ-field on $\mathbf{X}$.

Definition (Markov Kernel). *Given a nonempty set $\mathbf{X}$, and a σ -field $\mathcal{F}$ on $\mathbf{X}$, a Markov kernel is a function $P : \mathbf{X} \times \mathcal{F} \to [0,1]$ such that*

1. *$\forall E \in \mathcal{F}$, $P(\cdot, E)$ is measurable, and*
2. *$\forall x \in \mathbf{X}$, $P(x, \cdot)$ is a probability measure on $(\mathbf{X}, \mathcal{F})$.*

Definition (Stationary Measure/Distribution). *Given a measure space $(\mathbf{X}, \mathcal{F})$ and a Markov kernel P on that space, a probability measure $\pi : \mathcal{F} \to [0, \infty)$ is a stationary measure or P is stationary with respect to π, if $\forall E \in \mathcal{F}$,*

$$\int_{x \in \mathbf{X}} P(x, E)\pi(dx) = \pi(E).$$

Definition (Markov Operator). *Given a measure space $(\mathbf{X}, \mathcal{F})$ and a Markov kernel $P : \mathbf{X} \times \mathcal{F} \to [0,1]$ with stationary distribution $\pi : \mathcal{F} \to [0,1]$, we define the Markov Operator to be the function $\mathcal{P} : L^2(\pi) \to L^2(\pi)$ such that $\forall f \in L^2(\pi)$,*

$$\mathcal{P}f(x) := \int_{\mathbf{X}} f(y)P(x, dy), \qquad \forall x \in \mathbf{X}.$$

We denote $\overline{\mathbf{R}} := [-\infty, \infty]$ and $\overline{\mathbf{C}}$ similarly.

Theorem 1. *For any $f \in L^2(\pi)$, the Markov Operator evaluated at f, $\mathcal{P}f$, is $(\mathcal{F}, \mathcal{B}_{\overline{\mathbf{R}}})$ -measureable (equivalently $(\mathcal{F}, \mathcal{B}_{\overline{\mathbf{C}}})$ -measureable).*

Proof. As $f \in L^2(\pi)$, f is measureable and thus the integral $\int f(y)P(x, dy)$ is well-defined for every fixed $x \in \mathbf{X}$ (note that we haven't shown it is finite).

Notice that for every $x \in \mathbf{X}$, as

$$\int_{\mathbf{X}} f(y)P(x,dy) := \int_{\mathbf{X}} f^+(y)P(x,dy) - \int_{\mathbf{X}} f^-(y)P(x,dy),$$

where f^+ and f^- are the positive and negative parts of f, respectively, if $\int f^+(y)P(\cdot,dy)$ and $\int f^-(y)P(\cdot,dy)$ are measureable, so is $\int f(y)P(\cdot,dy) = \mathcal{P}f$. So we can assume without loss of generality that f is nonnegative. (A similar argument can be given for complex functions to only consider a positive real part).

As $f : \mathbf{X} \to \mathbf{R}$ is measureable, there exists a sequence of simple functions $\{\phi_n\}_{n\in\mathbf{N}}$, such that $0 \le |\phi_1| \le |\phi_2| \le \cdots \le |f|$, and $\phi_n \to f$ pointwise (Theorem 2.10, [2]). As f is nonnegative, we can assume $0 \le \phi_1 \le \phi_2 \le \cdots \le f$. As these are simple finctions, they have the form

$$\phi_n = \sum_{i=1}^{N_n} c_i^n \mathbf{1}_{E_i^n}, \quad \{c_i^n\} \subset \mathbf{R}, \quad \{E_i^n\} \subset \mathcal{F}, \quad N_n \in \mathbf{N}, \qquad \forall n \in \mathbf{N},$$

by definition. For every $n \in \mathbf{N}$, let $F_n : \mathbf{X} \to \mathbf{R}^\omega$ such that

$$F_n(x) = (c_1^n P(x,E_1^n), \dots, c_{N_n}^n P(x, E_{N_n}^n), 0, \dots), \qquad \forall x \in \mathbf{X},$$

and let $G : \mathbf{R}^\omega \to \mathbf{R}$ such that $G(x_1, x_2, \dots) = \sum_{i=1}^\infty x_i$, for every $(x_1, x_2, \dots) \in \mathbf{R}^\omega$. As $P(\cdot, E)$ is measureable for every $E \in \mathcal{F}$, $c_i^n P(\cdot, E_i^n)$ is measurebale for every $i \in \{1, \dots, N_n\}$ for every $n \in \mathbf{N}$. Thus F_n is measureable for every n by Proposition 2.4 of [2].

As G is continuous, it is also measureable (Corollary 2.2, [2]). So, $G \circ F_n : \mathbf{X} \to \mathbf{R}$ is measureable for every n. Furthermore,

$$(G \circ F_n)(x) = \sum_{i=1}^{N_n} c_i^n P(x, E_i^n) =: \int_{\mathbf{X}} \phi_n(y)P(x,dy), \qquad x \in \mathbf{X}.$$

As $\{\phi_n\}$ and f are nonnegative, $\phi_n \le \phi_{n+1}$ for every $n \in \mathbf{N}$, and $\phi_n \to f$ pointwise, by the Monotone Convergence Theorem (Theorem 2.14, [2]),

$$\mathcal{P}f(x) = \int f(y)P(x,dy) = \lim_{n\to\infty} \int \phi_n(y)P(x,dy) = \lim_{n\to\infty} (G \circ F_n)(x), \qquad \forall x \in \mathbf{X}.$$

Thus as $G \circ F_n$ is measureable for every n, $\mathcal{P}f$ is also measureable (Proposition 2.7, [2]). $\square$

NOTE: The above works for any $(\mathcal{F}, \mathcal{B}_{\overline{\mathbf{R}}})$ -measureable (equivalently $(\mathcal{F}, \mathcal{B}_{\overline{\mathbf{C}}})$ -measureable) f.

Theorem 2. *If P is a Markov kernel with stationary distribution π, if $f \in L^1(\pi)$, then $\pi(\mathcal{P}f) = \pi(f)$, i.e. $\mathbf{E}_\pi : L^1(\pi) \to \mathbf{R}$ (the expectation functional with respect to π) is invariant under $\mathcal{P}$.*

Proof. Let $f \in L^2(\pi)$. Then

$$\begin{aligned}
\mathbf{E}_\pi(\mathcal{P}f) &= \int_{x\in\mathbf{X}} \mathcal{P}f(x)\pi(dx) \\
&= \int_{x\in\mathbf{X}} \int_{y\in\mathbf{X}} f(y)P(x,dy)\pi(dx) \\
&= \int_{y\in\mathbf{X}} f(y) \int_{x\in\mathbf{X}} P(x,dy)\pi(dx) && \text{(Fubini's Theorem 2.37 in [2])} \\
&= \int_{y\in\mathbf{X}} f(y)\pi(dy) && \text{(as } P \text{ is stationary wrt } \pi) \\
&= \mathbf{E}_\pi(f).
\end{aligned}$$

□

Theorem 3. *If P is a Markov kernel with stationary distribution π, then for every $f \in L^2(\pi)$, $\mathcal{P}f \in L^2(\pi)$.*

Proof. This is just the case $r = 2$ in Lemma 1 from [4]. Following their proof,

$$\begin{aligned}
\|\mathcal{P}f\|^2_{L^2(\pi)} &= \int_{x\in\mathbf{X}} |\mathcal{P}f(x)|^2\pi(dx) \\
&= \int_{x\in\mathbf{X}} \left| \int_{y\in\mathbf{X}} f(y)P(x,dy) \right|^2 \pi(dx) \\
&\leq \int_{x\in\mathbf{X}} \left(\int_{y\in\mathbf{X}} |f(y)|P(x,dy) \right)^2 \pi(dx) && \text{(Triangle Inequality)} \\
&\leq \int_{x\in\mathbf{X}} \int_{y\in\mathbf{X}} |f(y)|^2 P(x,dy)\pi(dx) && \text{(Jensen's Inequality)} \\
&= \int_{y\in\mathbf{X}} |f(y)|^2 \int_{x\in\mathbf{X}} P(x,dy)\pi(dx) && \text{(Fubini's Theorem)} \\
&= \int_{y\in\mathbf{X}} |f(y)|^2\pi(dy) && \text{(as } P \text{ is stationary wrt } \pi) \\
&= \|f\|^2_{L^2(\pi)} \\
&< \infty. && \text{(as } f \in L^2(\pi))
\end{aligned}$$

□

2 Functional Analysis

Definition (Hilbert Spaces)**.** *A Hilbert space* $\mathbf{H}$ *is a linear space over* $\mathbf{R}$ *or* $\mathbf{C}$ *that is complete with respect to the norm generated by an inner product defined on* $\mathbf{H}$*.*

For the rest of this section, we assume $\mathbf{H}$ to be a Hilbert space, and the map $f \times g \mapsto \langle f, g \rangle$ from $\mathbf{H} \times \mathbf{H} \to \mathbf{C}$ to be the inner product defined on $\mathbf{H}$, such that $\mathbf{H}$ is complete with respect to the norm generated by this inner product, namely $\|f\| = \langle f, f \rangle^{1/2}$ for $f \in \mathbf{H}$.

Definition (Linear Operators)**.** *A function* $T : \mathbf{H} \to \mathbf{H}$ *is called an operator, and it is linear if for every* $\alpha, \beta \in \mathbf{C}$ *and for every* $f, g \in \mathbf{H}$*,*

$$T(\alpha f + \beta g) = \alpha T(f) + \beta T(g).$$

Definition (Bounded Operators)**.** *A linear operator* $T : \mathbf{H} \to \mathbf{H}$ *is bounded if there exists* $C > 0$ *such that for every* $f \in \mathbf{H}$*,*

$$\|Tf\| \leq C \|f\|.$$

We denote the space of bounded linear operators from $\mathbf{H} \to \mathbf{H}$ *by* $\mathfrak{B}(\mathbf{H})$*.*

Definition (Operator Norm)**.** *We define a norm on the operators, i.e. a function from* $\mathfrak{B}(\mathbf{H}) \to \mathbf{C}$*,* $T \mapsto \|T\|$ *as*

$$\|T\| := \sup\{\|Tf\| : f \in \mathbf{H}, \|f\| = 1\}.$$

Lemma 4. *Here we prove some equivalent definitions of the norm of an operator. For any* $T \in \mathfrak{B}(\mathbf{H})$*,*

$$\begin{aligned} \|T\| : &= \sup\{\|Tf\| : f \in \mathbf{H}, \|f\| = 1\} && (1) \\ &= \sup\left\{\frac{\|Tf\|}{\|f\|} : f \in \mathbf{H}, f \neq 0\right\} && (2) \\ &= \inf\{C > 0 : \|Tf\| \leq C\|f\|, \forall f \in \mathbf{H}\} && (3) \end{aligned}$$

Proof. We prove this by showing $(1) \leq (2) \leq (3) \leq (2) \leq (1)$. Let $T \in \mathfrak{B}(\mathbf{H})$.

$(1) \leq (2)$:

Notice if $f \in \mathbf{H}$ such that $\|f\| = 1$, then $\frac{\|Tf\|}{\|f\|} = \|Tf\|$. So,

$$\{\|Tf\| : f \in \mathbf{H}, \|f\| = 1\} \subset \left\{\frac{\|Tf\|}{\|f\|} : f \in \mathbf{H}, f \neq 0\right\}.$$

Thus taking the supremum of each set gives us our desired result.

$(2) \leq (3)$:

As T is bounded, there exists $C > 0$ such that $\|Tf\| \leq C \|f\|$ for every $f \in \mathbf{H}$. For any such C, notice for any $f \neq 0$,

$$\frac{\|Tf\|}{\|f\|} \leq \frac{C \|f\|}{\|f\|} = C.$$

Thus $\sup \left\{ \frac{\|Tf\|}{\|f\|} : f \in \mathbf{H}, f \neq 0 \right\} \leq \inf \{C > 0 : \|Tf\| \leq C \|f\|, \forall f \in \mathbf{H}\}$.

$(3) \leq (2)$:

Let $S := (2)$. Note that S exists as $\left\{ \frac{\|Tf\|}{\|f\|} : f \in \mathbf{H}, f \neq 0 \right\}$ is bounded from above, as T is a bounded operator.

Now for any $f \in \mathbf{H}$, $\|Tf\| \leq S \|f\|$, as if $f = 0$, then $\|T0\| = 0$. Thus, $S \in \{C > 0 : \|Tf\| \leq C \|f\|, \forall f \in \mathbf{H}\}$, and the result follows.

$(2) \leq (1)$:

For any $f \in \mathbf{H}$, $f \neq 0$, $g = \frac{f}{\|f\|} \in \mathbf{H}$ and $\|g\| = \frac{\|f\|}{\|f\|} = 1$, by the linearity of the norm. So, as T and the norm are linear,

$$\frac{\|Tf\|}{\|f\|} = \|Tg\|.$$

$\square$

Lemma 5. *Let $T : \mathbf{H} \to \mathbf{H}$ be a linear operator. The following statements are equivalent:*

1. *T is bounded. I.e. $T \in \mathfrak{B}(\mathbf{H})$.*
2. *T is continuous.*
3. *T is continuous at $0 \in \mathbf{H}$.*

Proof. $1 \implies 2$.

Let $f \in \mathbf{H}$ and let $\epsilon > 0$. Take $\delta = \epsilon / \|T\|$. Then for any $g \in \mathbf{H}$ such that $\|f - g\| < \delta$, as $T \in \mathfrak{B}(\mathbf{H})$,

$$\|Tf - Tg\| = \|T(f - g)\| \leq \|T\| \|f - g\| < \|T\| \delta = \epsilon.$$

$2 \implies 3$.

Trivial.

$3 \implies 1$.

Take $\epsilon = 1$. Then there exists $\delta > 0$ such that for every $f \in \mathbf{H}$ such that $\|f\| \leq \delta$, $\|Tf\| \leq 1$. Then for every $f \neq 0 \in \mathbf{H}$, let $\alpha_f = \frac{\delta}{\|f\|}$. Then $\|\alpha_f f\| = \alpha_f \|f\| = \delta$, so

$$\|Tf\| = \alpha_f^{-1} \|T(\alpha_f f)\| \leq \alpha_f^{-1} = \delta^{-1} \|f\|.$$

$\square$

Proposition 6. *If P is a Markov kernel with stationary distribution π, then $\mathcal{P} : L^2(\pi) \to L^2(\pi)$ is a bounded linear operator with operator norm equal to 1.*

Proof. $\mathcal{P}$ is clearly linear by the linearity of integrals, and we see from the proof of Theorem 3 that

$$\|\mathcal{P}f\| := \left(\int_{x\in\mathbf{X}} |\mathcal{P}f(x)|^2\pi(dx)\right)^{1/2} \leq \left(\int_{y\in\mathbf{X}} |f(y)|^2\pi(dy)\right)^{1/2} =: \|f\|,$$

so that in fact $\mathcal{P}$ is a bounded linear operator from $L^2(\pi)$ to itself.

For every $f \in L^2(\pi)$, by the above $\|\mathcal{P}f\| \leq \|f\|$. Thus $\|\mathcal{P}\| = \inf\{c > 0 : \|\mathcal{P}f\| \leq c\,\|f\|, f \in L^2(\pi)\} \leq 1$.

For the reverse inequality, notice that for $1 \in L^2(\pi)$, i.e. $1 : L^2(\pi) \to L^2(\pi)$ such that $1(x) = 1$ for every $x \in \mathbf{X}$, satisfies $\|1\| = 1$ as π is a probability distribution. Furthermore,

$$\|\mathcal{P}1\| = \int_{x\in\mathbf{X}} \left(\int_{y\in\mathbf{X}} P(x, dy)\right)^2 \pi(dx) = \int_{x\in\mathbf{X}} \pi(dx) = 1.$$

As $\|\mathcal{P}\| = \sup\{\|\mathcal{P}f\| : \|f\| = 1\}$, $\|\mathcal{P}\| \geq \|\mathcal{P}1\| = 1$. □

For what follows, unless explicitly stated otherwise, treat $\mathcal{P}$ as an operator from $L^2(\pi)$ to itself.

Definition (Adjoints)**.** *If $T \in \mathfrak{B}(\mathbf{H})$, then we define $T^* \in \mathfrak{B}(\mathbf{H})$ such that*

$$\langle Tf, g\rangle = \langle f, T^*g\rangle \quad \forall f, g \in \mathbf{H},$$

and $\|T\| = \|T^*\|$ to be the adjoint of T.

Theorem 7. *For every bounded linear operator $T \in \mathfrak{B}(\mathbf{H})$, there exists a unique adjoint of T, $T^* \in \mathfrak{B}(\mathbf{H})$.*

Proof. Refer to [7] Theorem 12.8 and 12.9. □

Definition (Self-Adjoint and Normal Operators)**.** *An operator $S \in \mathfrak{B}(\mathbf{H})$ is called self-adjoint if $S = S^*$, i.e. S equals it's adjoint.*

An operator $N \in \mathfrak{B}(\mathbf{H})$ is normal if $NN^* = N^*N$, i.e. N commutes with it's adjoint.

It is obvious from the definitions that self-adjoint operators are also normal operators.

Definition (Reversibility)**.** *We say that a Markov Kernel P is reversible with respect to a probability measure π if*

$$P(x, dy)\pi(dx) = P(y, dx)\pi(dy), \qquad \pi\text{-a.e. } x, y \in \mathbf{X}.$$

Proposition 8. *If P is a reversible Markov kernel with respect to the probability measure π, then π is a stationary measure of P.*

Proof. For $x, y \in \mathbf{X}$, as $P(y, \cdot)$ is a probability measure for every y, we have

$$\int_{x \in \mathbf{X}} P(x, dy)\pi(dx) = \int_{x \in \mathbf{X}} P(y, dx)\pi(dy) = \pi(dy).$$

$\square$

Theorem 9. *If P is a Markov kernel with stationary distribution π, then P is reversible with respect to π if and only if $\mathcal{P} : L^2(\pi) \to L^2(\pi)$ is self-adjoint.*

Proof. Let $f, g \in L^2(\pi)$. Then

$$\begin{aligned}
\langle \mathcal{P} f, g \rangle &= \int_{x \in \mathbf{X}} (\mathcal{P} f)(x) g(x) \pi(dx) \\
&= \int_{x \in \mathbf{X}} \left(\int_{y \in \mathbf{X}} f(y) P(x, dy) \right) g(x) \pi(dx) \\
&= \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(y) g(x) P(x, dy) \pi(dx) \\
&= \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(y) g(x) P(y, dx) \pi(dy) && \text{(by reversibility of } P\text{)} \\
&= \int_{y \in \mathbf{X}} \int_{x \in \mathbf{X}} f(y) g(x) P(y, dx) \pi(dy) && \text{(by Fubini's Theorem)} \\
&= \int_{y \in \mathbf{X}} f(y) \left(\int_{x \in \mathbf{X}} g(x) P(y, dx) \right) \pi(dy) \\
&= \int_{y \in \mathbf{X}} f(y) (\mathcal{P} g)(y) \pi(dy) \\
&= \langle f, \mathcal{P} g \rangle.
\end{aligned}$$

So, as $f, g \in L^2(\pi)$ are arbitrary, $\langle f, \mathcal{P} g \rangle = \langle f, \mathcal{P}^* g \rangle$ for every $f, g \in L^2(\pi)$, so $\mathcal{P} = \mathcal{P}^*$.

For the converse, say $\mathcal{P}$ is self-adjoint. Then for any $f, g \in L^2(\pi)$,

$$\begin{aligned}
&\int_{x\in\mathbf{X}}\int_{y\in\mathbf{X}} f(y)g(x)P(x,dy)\pi(dx) \\
&= \int_{x\in\mathbf{X}} (\mathcal{P}f)(x)g(x)\pi(dx) \\
&= \langle \mathcal{P}f, g\rangle \\
&= \langle f, \mathcal{P}g\rangle. && \text{(as } \mathcal{P} \text{ is self-adjoint)} \\
&= \int_{y\in\mathbf{X}} f(y)(\mathcal{P}g)(y)\pi(dy) \\
&= \int_{y\in\mathbf{X}}\int_{x\in\mathbf{X}} f(y)g(x)P(y,dx)\pi(dy) \\
&= \int_{x\in\mathbf{X}}\int_{y\in\mathbf{X}} f(y)g(x)P(y,dx)\pi(dy). && \text{(by Fubini's Theorem)}
\end{aligned}$$

Thus by definition, for $x, y \in \mathbf{X}$, $P(x,dy)\pi(dx) = P(y,dx)\pi(dy)$. □

It's worth noting here that according to Mira and Geyer in [5] on page 7, there are a lot of Markov chains whose Markov operator is not compact. This is why we must move to a more generalized spectrum. It may be worth exploring some of these examples to show. They cite a source.

Definition (Invertible Bounded Operators)**.** *A bounded linear operator $T \in \mathfrak{B}(\mathbf{H})$ is called invertible if T is bijective and it's inverse $T^{-1} : \mathbf{H} \to \mathbf{H}$ defined as $T^{-1}f = g$ where $Tg = f$, for every $f \in \mathbf{H}$, is also bounded. I.e. $T^{-1} \in \mathfrak{B}(\mathbf{H})$.*

Note that this can be a large point of confusion, as the usual notion of inverse is simply that the inverse map exists and is well-defined. However here, when we talk about the inverse of a bounded operator, we not only mean that the inverse map exists and is well-defined, but that is ALSO bounded. Similar to how a homeomorphism between spaces is a continuous map between topological spaces that is bijective and it's inverse is also continuous, an invertible operator is a bounded linear operator between Hilbert spaces (or more generally vector spaces) that is bijective and it's inverse is also a bounded linear operator. (Note that it is trivial if a bounded linear operator is bijective it's inverse is linear. The inverse being bounded is not trivial, and requires more than being bijective).

Definition (Spectrum)**.** *For $T \in \mathfrak{B}(\mathbf{H})$, we call $\sigma(T) \subset \mathbf{C}$ such that*

$$\sigma(T) := \{\lambda \in \mathbf{C} : T - \lambda \textit{ is not invertible}\},$$

the spectrum of T.

Theorem 10. *If T is a bounded linear operator on a Hilbert space $\mathbf{H}$ ($T \in \mathfrak{B}(\mathbf{H})$), then*

$$\sup\{|\lambda| : \lambda \in \sigma(T)\} = \lim_{n\to\infty} \|T^n\|^{1/n}.$$

Proof. Relies on some complex analysis. Proposition 3.8 of chapter VII in Conway [6]. □

Lemma 11. *If T is a bounded linear operator on a Hilbert space $\mathbf{H}$ ($T \in \mathfrak{B}(\mathbf{H})$), then for every $\lambda \in \sigma(T)$, $|\lambda| \leq \|T\|$.*

Proof. Let $n \in \mathbf{N}$, and let $f \in \mathbf{H}$ such that $\|f\| = 1$. Then

$$\|T^n f\| \leq \|T\|^n \|f\| = \|T\|^n.$$

So by the first representation of the operator norm of lemma 4, $\|T^n\|^{1/n} \leq \|T\|$ for every $n \in \mathbf{N}$.

So, by theorem 10,

$$\sup\{|\lambda| : \lambda \in \sigma(T)\} = \lim_{n\to\infty} \|T^n\|^{1/n} \leq \|T\|.$$

□

It is important to note the difference here between finite dimensional vector spaces and general Hilbert spaces.

Theorem 12. *If $\mathbf{V}$ is a finite dimensional Hilbert space and $T : \mathbf{V} \to \mathbf{V}$ is linear, then T is bounded, $T \in \mathfrak{B}(\mathbf{V})$. In particular, if $T \in \mathfrak{B}(\mathbf{V})$ is injective, then T is invertible.*

Proof. As $\mathbf{V}$ is finite dimensional, let n be the dimension of $\mathbf{V}$. Then there exists an orthonormal basis for $\mathbf{V}$, $\{e_i\}_{i=1}^n \subset \mathbf{V}$. Let $M = \max\{\|Te_i\| : i \in \{1, \dots, n\}\}$.

Let $\epsilon > 0$, and take $\delta = \epsilon/(2Mn)$. Let $v \in \mathbf{V}$ such that $\|v\| \leq \delta$. Then as $v \in \mathbf{V}$, there exists $\alpha_1, \dots, \alpha_n \in \mathbf{C}$ such that $v = \alpha_1 e_0 + \dots + \alpha_n e_n$. Let $m_v = \max\{|\alpha_i| : i \in \{1, \dots, n\}\}$. As T is linear,

$$\begin{aligned}
\|Tv\| &= \|T(\alpha_0 e_0 + \dots + \alpha_n e_n)\| \\
&\leq |\alpha_0| \|Te_0\| + \dots |\alpha_n| \|Te_n\| \\
&\leq M \sum_{i=0}^{n} |\alpha_i| \\
&\leq Mnm_v.
\end{aligned}$$

As $\|v\| \leq \delta$, we have $|\alpha_i| \leq \|v\| \leq \delta$ for every $i \in \{1, \ldots, n\}$, and thus $m_v \leq \delta$. So, we have

$$\|Tv\| \leq M \sum_{i=1}^{n} |\alpha_i| \leq Mnm_v \leq Mn\delta \leq \epsilon/2 < \epsilon.$$

So, T is continuous at $0 \in \mathbf{V}$, and thus by Lemma 5, T is bounded.

So, if $T \in \mathfrak{B}(\mathbf{V})$ is injective, then notice $\{Te_i\}_{i=1}^n \subset \mathbf{V}$ is a linearly independent set of vectors, as otherwise, if $Te_j = \sum_{i \neq j} T\alpha_i e_i$ for some $\{\alpha_i\} \subset \mathbf{C}$, we have

$$0 = Te_j - \sum_{i \neq j} T\alpha_i e_i = T(e_j - \sum_{i \neq j} \alpha_i e_i).$$

As T is injective and $T0 = 0$, we have $e_j - \sum_{i \neq j} \alpha_i e_i = 0$, i.e. $e_j = \sum_{i \neq j} \alpha_i e_i$, which contradicts $\{e_i\}_{i=1}^n$ being orthonormal.

So, as $\{Te_i\}_{i=1}^n$ is a set of n linearly independent vectors and $\mathbf{V}$ has dimension n, $\mathbf{V} = \text{span}\{Te_i\}_{i=1}^n$, and thus T is also surjective.

So, T is bijective, and thus it's inverse T^{-1} exists and is also linear, and thus by the first part of this theorem is also bounded. □

This theorem highlights how in finite dimensions, the set of eigenvalues is exactly the spectrum of the operator. In the general case however, things are not so simple.

For $E \subset \mathbf{H}$, we define $E^{\perp} \subset \mathbf{H}$ to be

$$E^{\perp} := \{f \in \mathbf{H} | \langle f, e \rangle = 0, \quad \forall e \in E\}.$$

Proposition 13. *For every bounded operator T on a Hilbert space $\mathbf{H}$,* $\text{null}(T^*) = \text{range}(T)^{\perp}$.

Proof. $f \in \mathbf{H}$ exists in $\text{null}(T^*)$ if and only if $T^* f = 0$, if and only if $\langle T^* f, h \rangle = 0$ for every $h \in \mathbf{H}$, if and only if $\langle f, Th \rangle = 0$ for every $h \in \mathbf{H}$, if and only if $f \in \text{range}(T)^{\perp}$. □

Proposition 14. *If T is a normal operator on the Hilbert space $\mathbf{H}$, then $\|Tf\| = \|T^* f\|$ for every $f \in \mathbf{H}$. Furthermore, T is injective if and only if T^* is injective, so* $\text{range}(T)$ *is dense in $\mathbf{H}$ if and only if T is injective.*

Proof. Let $T \in \mathfrak{B}(\mathbf{H})$ be normal. Then

$$\begin{aligned}
\|Tf\| &= \langle Tf, Tf \rangle^{1/2} \\
&= \langle T^*Tf, f \rangle^{1/2} \\
&= \langle TT^*f, f \rangle^{1/2} && \text{(as } T \text{ is normal)} \\
&= \langle T^*f, T^*f \rangle^{1/2} \\
&= \|T^*f\|.
\end{aligned}$$

Say T is injective. Then T^* is injective, as if not, there exists $f \in \mathbf{H}, f \neq 0$, such that $T^* f = 0$. But then $0 = \|T^* f\| = \|Tf\|$, so $Tf = 0$, which can't happen because T is injective. Similarly, if T^* is injective, then so is T.

Say T is injective. Then T^* is injective, so

$$\operatorname{range}(T)^\perp = \operatorname{null}(T^*) = \{0\},$$

and thus $\left(\operatorname{range}(T)^\perp\right)^\perp = \{0\}^\perp = \mathbf{H}$.

Say $\operatorname{range}(T)$ is dense in $\mathbf{H}$. Then

$$\{0\} = \operatorname{range}(T)^\perp = \operatorname{null}(T^*),$$

so T^* is injective, so T is injective also. □

Proposition 15. *Say (X, d_1) and (Y, d_2) metric spaces with (X, d_1) complete. Then for any continuous function $f : X \to Y$, if*

$$d_2(f(x), f(x')) \geq \alpha d_1(x, x')$$

for every $x, x' \in X$ for some $\alpha > 0$, then $\operatorname{range}(f)$ *is closed.*

Proof. Let $y \in \overline{\operatorname{range}(f)}$. Then there exists $(x_n) \subset X$ such that $\lim_{n\to\infty} f(x_n) = y$. Thus $(f(x_n))$ is a Cauchy sequence in Y.

We now show (x_n) is a Cauchy sequence in X. Let $\epsilon > 0$. As $(f(x_n))$ is a Cauchy sequence, pick $N \in \mathbf{N}$ such that for every $n, m \geq N$, $d_2(f(x_n, x_m)) < \epsilon/\alpha$. Then for every $n, m \geq N$,

$$d_1(x_n, x_m) \leq \frac{d_2(f(x_n, x_m))}{\alpha} < \epsilon.$$

As X is complete, (x_n) converges in X, say to $x \in X$. Then as f is continuous, $f(x) = \lim_{n\to\infty} f(x_n) = y$, so $y \in \operatorname{range}(f)$. □

Lemma 16. *If T is a normal operator, then T is invertible if there exists $\delta > 0$ such that $\|Tf\| \geq \delta \|f\|$ for every $f \in \mathbf{H}$.*

Note that this condition is actually an if and only if, though the other direction uses the open mapping lemma, and isn't required in the proof that the spectrum of a self-adjoint operator is real.

Proof. First we notice that given the hypothesis, T is injective, as for every $f \in \mathbf{H}$ such that $f \neq 0$, $\|Tf\| \geq \delta \|f\| > 0$, so $Tf \neq 0$. Thus, by Proposition 14, $\operatorname{range}(T)$ is dense in $\mathbf{H}$. Next notice that by Proposition 15, $\operatorname{range}(T)$ is closed in $\mathbf{H}$. So,

$$\mathbf{H} = \overline{\operatorname{range}(T)} = \operatorname{range}(T).$$

So, T is invertible. □

Lemma 17. *If T is a normal operator, then $\sigma(T) \subset \mathbf{R}$.*

Proof. Let $\lambda \in \sigma(T)$. Then write $\lambda = \alpha + i\beta$, and assume for a contradiction that $\beta \neq 0$. Then for every $f \in \mathbf{H}$,

$$\|(T-\lambda)f\|^2 = \|Tf - \alpha f - i\beta f\|^2 = \|Tf - \alpha f\|^2 + |\beta|^2 \|f\|^2,$$

and so $\|(T-\lambda)f\| \geq |\beta| \; \|f\|$.

So by Lemma 16, $T - \lambda$ is invertible, and thus $\lambda \notin \sigma(T)$. □

Proposition 18. *If P is a reversible Markov kernel with respect to π, then the spectrum of the Markov operator as a function from $L^2(\pi)$ to $L^2(\pi)$, is a subset of $[-1,1]$. That is for $\mathcal{P} : L^2(\pi) \to L^2(\pi)$, $\sigma(\mathcal{P}) \subset [-1,1]$.*

Proof. As P is reversible, by lemma 9 $\mathcal{P}$ is self-adjoint and thus also normal, so by lemma 17, $\sigma(\mathcal{P}) \subset \mathbf{R}$.

By lemma 11, for every $\lambda \in \sigma(\mathcal{P})$, $|\lambda| \leq \|\mathcal{P}\| = 1$ by proposition 6, as π is a stationary distribution of P by proposition 8. Thus, as $\sigma(\mathcal{P}) \subset \mathbf{R}$, we have $\sigma(\mathcal{P}) \subset [-1,1]$. □

Definition (Projections)**.** *An operator $T \in \mathfrak{B}(\mathbf{H})$ is called a projection if $T^2 = T$.*

Definition (Spectral Measure)**.** *(From Conway [6] page 256. Mira Geyer[5] call the spectral measure the next definition, and this measure the resolution of the identity. That is also what Rudin [7] calls it). Given a set a measurable space $(Y, \mathcal{M})$ and a Hilbert space $\mathbf{H}$, a spectral measure on $(Y, \mathcal{M}, \mathbf{H})$ is a function $\mathcal{E} : \mathcal{M} \to \mathfrak{B}(\mathbf{H})$ such that*

1. *for every $A \in \mathcal{M}$, $\mathcal{E}(A)$ is a self-adjoint projection,*
2. *$\mathcal{E}(\emptyset) = 0$ and $\mathcal{E}(Y) = 1$,*
3. *for every $A, B \in \mathcal{M}$, $\mathcal{E}(A \cap B) = \mathcal{E}(A)\mathcal{E}(B)$,*
4. *if $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ are pairwise disjoint, then $\mathcal{E}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathcal{E}(A_n)$.*

Definition (Eigenmeasure)**.** *Given a spectral measure $\mathcal{E}$ on the space $(Y, \mathcal{M}, \mathbf{H})$, an Eigenmeasure with respect to $f, g \in \mathbf{H}$ is a function $E_{f,g} : \mathcal{M} \to \mathbf{C}$ such that*

$$E_{f,g}(B) := \langle \mathcal{E}(B)f, g \rangle, \quad \forall B \in \mathcal{M}.$$

Theorem 19 (Spectral Theorem)**.** *If T is a normal bounded linear operator on the Hilbert space $\mathbf{H}$, then there is a unique spectral measure $\mathcal{E}$ on the Borel subsets of $\sigma(T)$ such that*

$$T = \int_{\lambda \in \sigma(T)} \lambda \mathcal{E}(d\lambda).$$

Remark. *For a discussion on the above spectral theory, see Rudin [7] chapter 12, or Conway [6] chapter IX.2.*

In our context, given a reversible Markov kernel P, $\mathcal{P}$ is self-adjoint and thus also normal, so the spectral Theorem applies. So, $\mathcal{P}$ has a spectral decomposition, and for every $f \in L^2(\pi)$, there exists an associated eigenmeasure. We denote the spectral measure of a Markov operator $\mathcal{P}$ by $\mathcal{E}_{\mathcal{P}}$, or just $\mathcal{E}$ when $\mathcal{P}$ is understood. For our purposes, as the eigenmeasure will almost always be for an inner product of the same function. For every $f \in L^2(\pi)$, we express the eigenmeasure of $\langle f, \mathcal{P}f\rangle$ as just $E_{\mathcal{P}}$, so

$$\langle f, \mathcal{P}f\rangle = \int_{\lambda \in \sigma(\mathcal{P})} \lambda E_{\mathcal{P}}(d\lambda),$$

where $E_{\mathcal{P}}(\cdot) := \langle f, \mathcal{E}(\cdot)f\rangle$, where $\mathcal{E} : \mathcal{B}_{\sigma(\mathcal{P})} \to \mathfrak{B}(L^2(\pi))$ is the unique spectral measure of $\mathcal{P}$ implied by the spectral Theorem. In the case where we have more than just one function in the inner product, we express the eigenmeasure as follows. For $f, g \in L^2(\pi)$

$$E_{f,g,\mathcal{P}}(\cdot) := \langle f, \mathcal{E}(\cdot)g\rangle.$$

An important fact about the eigenmeasure of the same function on both sides of the inner product is the following.

Lemma 20. *For any normal bounded linear operator T on the hilbert space $\mathbf{H}$, for any $f \in \mathbf{H}$, the eigenmeasure of T with respect to f, E_T, is a non-negative measure.*

Proof. Let $A \in \mathcal{B}(\mathbf{C})$. Then

$$\begin{aligned} E_T(A) :&= \langle f, \mathcal{E}_T(A)f\rangle \\ &= \langle f, \mathcal{E}_T(A)^2 f\rangle && \text{(as } \mathcal{E}_T \text{ is a projection)} \\ &= \langle \mathcal{E}_T(A)f, \mathcal{E}_T(A)f\rangle && \text{(as } \mathcal{E}_T \text{ is self-adjoint)} \\ &=: \|\mathcal{E}_T(A)f\|^2 \geq 0. \end{aligned}$$

□

Definition (Point-Spectrum)**.** *We call $\sigma_p(T) \subset \mathbf{C}$ such that*

$$\sigma_p(T) := \{\lambda \in \mathbf{C} : \mathbf{H} \supset \ker(T - \lambda) \neq \{0\}\},$$

the point spectrum of T.

Notice that $\sigma_p(T)$ is exactly the set of eigenvalues of T, and $\ker(T - \lambda)$ is exactly the set of eigenfunctions of T with eigenvalue $\lambda \in \mathbf{C}$.

We denote by $\mathbf{H}_{\mathbf{R}}$ the space of $f \in \mathbf{H}$ such that the coefficients of f are real.

Lemma 21. *If T is a self-adjoint operator on $\mathbf{H}$, then*

1. *the eigenvalues of T are real,*
2. *the eigenvalues of T have real eigenfunctions,*
3. *and eigenfunctions with distinct eigenvalues are orthogonal.*

Proof. 1. Let λ be an eigenvalue of T. Then there exists $u \neq 0 \in \mathbf{H}$ such that $Tu = \lambda u$. So,

$$\lambda \left\|u\right\|^2 = \lambda\langle u, u\rangle = \langle Tu, u\rangle = \langle u, Tu\rangle = \overline{\lambda}\langle u, u\rangle = \overline{\lambda} \left\|u\right\|^2 .$$

So $(\lambda - \overline{\lambda}) \left\|u\right\|^2 = 0$. As $u \neq 0$, $\left\|u\right\|^2 \neq 0$, thus $\lambda = \overline{\lambda}$, so $\lambda \in \mathbf{R}$.

2. Let λ be an eigenvalue of T. Then there exists $u \neq 0 \in \mathbf{H}$ such that $Tu = \lambda u$. As $u \in \mathbf{H}$, there exists $a, b \in \mathbf{H_R}$ such that $u = a + ib$. So,

$$Ta + iTb = T(a + ib) = Tu = \lambda u = \lambda a + i\lambda b.$$

Furthermore, as $u \neq 0$, $a \neq 0$ or $b \neq 0$. So either $a \neq 0$ or $b \neq 0$, and is thus a real eigenfunction of T with eigenvalue λ.

3. Let λ_1, λ_2 be distinct eigenvalues of T with eigenfunctions u_1 and u_2 respectively. Then as T is self-adjoint and $\lambda_1, \lambda_2 \in \mathbf{R}$ by (1), so

$$(\lambda_1 - \lambda_2)\langle u_1, u_2\rangle = \langle Tu_1, u_2\rangle - \langle u_1, Tu_2\rangle = \langle Tu_1, u_2\rangle - \langle Tu_1, u_2\rangle = 0.$$

As λ_1 and λ_2 are disjoint, $(\lambda_1 - \lambda_2) \neq 0$, so $\langle u_1, u_2\rangle = 0$. □

Proposition 22. *If P is a reversible Markov kernel, then*

1. *the eigenvalues of $\mathcal{P}$ are real,*
2. *the eigenvalues of $\mathcal{P}$ have real eigenfunctions,*
3. *and eigenfunctions with distinct eigenvalues are orthogonal.*

Proof. As P is reversible, by lemma 9, $\mathcal{P}$ is self-adjoint. Lemma 21 completes it. □

We now discuss the eigenvalue 1 of a Markov operator.

Proposition 23. *If P is a Markov kernel (not necessarily reversible), then 1 is an eigenvalue of $\mathcal{P}$. Furthermore, any π-almost everywhere constant function is an eigenfunction of $\mathcal{P}$ with eigenvalue 1, where π is a stationary distribution of P.*

Proof. Let $c : \mathbf{X} \to \mathbf{R}$ (or $\mathbf{C}$) where $c \equiv K \in \mathbf{R}$ (or $\mathbf{C}$), π-almost everywhere. Then for any $x \in \mathbf{X}$, (as we are considering $c \in L^2(\pi)$, we can take $c \equiv K$ everywhere),

$$\mathcal{P}c(x) := \int_{y \in \mathbf{X}} KP(y, dx) = K = c(x).$$

□

Definition. *We now define the subspace*

$$L_0^2(\pi) := \{f \in L^2(\pi) : f \perp 1\} \subset L^2(\pi),$$

where $1 \in L^2(\pi)$ *is the function identically equal to* 1 π*-almost everywhere.*

Remark. *Another more probibalistic interpretation is that* $L_0^2(\pi) = \{f \in L^2(\pi) : \mathbf{E}_\pi(f) = 0\}$*, as*

$$\mathbf{E}_\pi(f) := \int_{x \in \mathbf{X}} f(x)\pi(dx) = \langle f, 1 \rangle.$$

Lemma 24. *If* P *is a Markov kernel (not necessarily reversible) with* π *as a stationary distribution, then* $\mathcal{P}$ *restricted to* $L_0^2(\pi)$*, maps back to* $L_0^2(\pi)$*. I.e.* $\mathrm{range}(\mathcal{P}|_{L_0^2(\pi)}) \subset L_0^2(\pi)$*.*

Proof. Let $f \in L_0^2(\pi)$. Then

$$\begin{aligned}
\langle \mathcal{P}f, 1 \rangle :=& \int_{x \in \mathbf{X}} \mathcal{P}f(x)\pi(dx) \\
=& \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(y)P(x, dy)\pi(dx) \\
=& \int_{y \in \mathbf{X}} f(y) \int_{x \in \mathbf{X}} P(x, dy)\pi(dx) && \text{(by Fubini's Theorem)} \\
=& \int_{y \in \mathbf{X}} f(y)\pi(dy) && \text{(as } P \text{ is stationary wrt } \pi) \\
=& \langle f, 1 \rangle \\
=& 0. && \text{(as } f \in L_0^2(\pi))
\end{aligned}$$

So $\mathcal{P}f \in L_0^2(\pi)$. □

We just showed that for a Markov kernel P with stationary distribution π, $\mathcal{P}$ is a well-defined map from $L_0^2(\pi) \to L_0^2(\pi)$. From here on out, unless stated otherwise, we restrict ourselves to the subspace $L_0^2(\pi)$ of $L^2(\pi)$.

Definition (φ-Irreducibility)**.** *A Markov kernel* P *on* $(\mathbf{X}, \mathcal{F})$ *is* φ*-irreducible if there exists a non-zero* σ*-finite measure* φ *on* $(\mathbf{X}, \mathcal{F})$ *such that for every* $E \in \mathcal{F}$ *such that* $\varphi(E) > 0$*, for every* $x \in \mathbf{X}$ *there exists* $n \in \mathbf{N}$ *such that* $P^n(x, E) > 0$*.*

The reason we restrict ourselves to this subspace, is on this subspace, if P is φ-irreducible, then 1 is not an eigenvalue of $\mathcal{P}$ on $L^2_0(\pi)$. We prove this result here.

Lemma 25. *If P is a φ-irreducible (not necessarily reversible) Markov kernel with stationary distribution π, then 1 is NOT an eigenvalue of $\mathcal{P}$ when restricted to $L^2_0(\pi)$.*

Proof. We provide only a very rough sketch of this proof, as the full proof is very involved, and uses facts about martingales, a stochastic process not included in this paper. We follow the same proof as in [11].

As P is φ-irreducible with stationary distribution π, it follows that π is the unique stationary distribution for P. (As by Proposition 10.1.1 of [12] P is recurrent, so Theorem 10.0.1 of [12] proves it).

We then see that if $f \in L^2(\pi)$ is an eigenvector of $\mathcal{P}$ with eigenvalue 1, then we must have $f(x) = \mathbf{E}_\pi(f)$ for π-almost every $x \in \mathbf{X}$, and thus f is a constant function. As f is an eigenvector, $f(x) \neq 0$ for π-almost every $x \in \mathbf{X}$, so $\mathbf{E}_\pi(f) \neq 0$. And thus $f \notin L^2_0(\pi)$, as $\langle f, 1\rangle = \mathbf{E}_\pi(f) \neq 0$.

(The above is a rough sketch of the proof of Proposition 22.1.2 of [13]. The full proof uses the fact that for every $n \in \mathbf{N}$, where $X_n : \mathbf{X} \to \mathbf{X}$ is random variable that equals the state space of the Markov chain with Markov kernel P at time n started from the stationary distribution π,

$$\mathbf{E}_\pi(f(X_n)) = \int_{x\in\mathbf{X}} \mathcal{P}^n f(x)\pi(dx) = \mathbf{E}_\pi(f),$$

so $\{f(X_n) : n \in \mathbf{N}\}$ is a martingale.) □

Remark. *Note that the above is NOT saying that $1 \notin \sigma(\mathcal{P})$ when $\mathcal{P} : L^2_0(\pi) \to L^2_0(\pi)$, but that 1 is not an eigenvalue, so there doesn't exist $f \in L^2_0(\pi)$ such that $\mathcal{P}f = f$.*

Lemma 26. *If $T \in \mathfrak{B}(\mathbf{H})$ is a normal bounded linear operator, with spectral decomposition $T = \int_{\lambda\in\sigma(T)} \lambda\mathcal{E}_T(d\lambda)$, then if $\lambda \in \sigma(T)$ is not an eigenvalue of T, then for every $f \in \mathbf{H}$, $E_{f,T}(\{\lambda\}) = 0$.*

Proof. By Theorem 12.29 (b) of [7], as $\lambda \in \sigma(T)$ is not an eigenvalue of T, $\mathcal{E}_T(\{\lambda\}) = 0$. So for every $f \in \mathbf{H}$,

$$E_{f,T}(\{\lambda\}) := \langle f, \mathcal{E}_T(\{\lambda\})f\rangle = \langle f, 0\rangle = 0.$$

□

Proposition 27. *If P is a φ-irreducible (not necessarily reversible) Markov kernel with stationary distribution π, then for every $f \in L^2_0(\pi)$, $E_{\mathcal{P}}(\{1\}) = 0$.*

Proof. By lemma 25, 1 is not an eigenvalue of $\mathcal{P}$. So if $1 \in \sigma(\mathcal{P})$, then by lemma 26 for any $f \in L_0^2(\pi)$, $E_{\mathcal{P}}(\{1\}) = 0$. If $1 \notin \sigma(\mathcal{P})$ then for every $f \in L_0^2(\pi)$, $E_{\mathcal{P}}(\{1\}) = 0$ trivially. □

(From [6] page 264.) We now define what we mean when we take a function of an operator. For every Borel measureable bounded function $\phi : \sigma(\mathcal{P}) \to \mathbf{C}$ where $\mathcal{P}$ is a self adjoint operator,

$$\phi(\mathcal{P}) := \int_{\lambda \in \sigma(\mathcal{P})} \phi(\lambda) \mathcal{E}_{\mathcal{P}}(d\lambda),$$

where $\mathcal{P} = \int_{\lambda \in \sigma(\mathcal{P})} \lambda \mathcal{E}_{\mathcal{P}}(d\lambda)$ is the spectral decomposition of $\mathcal{P}$. As a consequence

$$\langle f, \phi(\mathcal{P}) f \rangle = \int_{\lambda \in \sigma(\mathcal{P})} \phi(\lambda) E_{\mathcal{P}}(d\lambda).$$

An important function is $\mathcal{P} \mapsto \mathcal{P}^k$ for nonnegative k. Notice that as $\sigma(\mathcal{P}) \subset [-1, 1]$, as $[-1, 1]$ is compact and $\cdot^k$ is continuous, it is bounded by it's maximum and minimum. It is also Borel measureable being continuous. Furthermore, notice for any $x \in \mathbf{X}$,

$$\begin{aligned}
(\mathcal{P}^k f)(x) &= \mathcal{P}^{k-1} \circ (\mathcal{P} f)(x) \\
&= \mathcal{P}^{k-1} \circ \int_{x_1 \in \mathbf{X}} f(x_1) P(x, dx_1) \\
&= \mathcal{P}^{k-2} \circ \mathcal{P} \left(\int_{x_1 \in \mathbf{X}} f(x_1) P(x, dx_1) \right) \\
&= \mathcal{P}^{k-2} \circ \int_{x_2 \in \mathbf{X}} \int_{x_1 \in \mathbf{X}} f(x_1) P(x_2, dx_1) P(x, dx_2) \\
&\vdots \\
&= \int_{x_k \in \mathbf{X}} \cdots \int_{x_1 \in \mathbf{X}} f(x_1) P(x_2, dx_1) \cdots P(x, dx_k) \\
&=: \int_{y \in \mathbf{X}} f(y) P^k(x, dy).
\end{aligned}$$

3 Relation to Asymptotic Variance

Definition (Asymptotic Variance)**.** *Given a Markov kernel P with stationary distribution π and $f \in L^2(\pi)$, we define the asymptotic variance as*

$$v(f, P) := \lim_{N \to \infty} \left[\frac{1}{N} \mathbf{Var} \left(\sum_{n=1}^{N} f(X_n) \right) \right],$$

where $\{X_n\}_{n=0}^{\infty}$ is the Markov chain with Markov kernel P started from stationarity (X_0 chosen wrt π).

Lemma 28. *For every $f \in L_0^2(\pi)$ and for every $N \in \mathbf{N}$,*

$$\frac{1}{N}\mathbf{Var}\left(\sum_{n=1}^{N} f(X_n)\right) = \langle f, f\rangle + 2\sum_{k=1}^{N}\left(\frac{N-k}{N}\right)\langle f, \mathcal{P}^k f\rangle.$$

Proof. Notice that

$$\frac{1}{N}\mathbf{Var}\left(\sum_{n=1}^{N} f(X_n)\right) = \frac{1}{N}\mathbf{E}_{\pi,P}\left[\left(\sum_{n=1}^{N} f(X_n)\right)^2\right] - \frac{1}{N}\mathbf{E}_{\pi,P}\left[\sum_{n=1}^{N} f(X_n)\right]^2.$$

But by the linearity of $\mathbf{E}$, we have

$$\begin{aligned}
\mathbf{E}_{\pi,P}\left[\sum_{n=1}^{N} f(X_n)\right] &= \sum_{n=1}^{N}\mathbf{E}_{\pi,P}[f(X_n)] \\
&= \sum_{n=1}^{N}\mathbf{E}_{\pi}[f] \\
&= 0. && (\text{as } f \in L_0^2(\pi))
\end{aligned}$$

Now by expanding the square and using the linearity of $\mathbf{E}$,

$$\begin{aligned}
\mathbf{E}_{\pi,P}\left[\left(\sum_{n=1}^{N} f(X_n)\right)^2\right] &= \mathbf{E}_{\pi,P}\left[\left(\sum_{n=1}^{N} f(X_n)^2\right) + 2\sum_{n=1}^{N}\sum_{i=1}^{k-1} f(X_n)f(X_i)\right] \\
&= \sum_{n=1}^{N}\mathbf{E}_{\pi,P}[f(X_n)^2] + 2\sum_{n=1}^{N}\sum_{i=1}^{k-1}\mathbf{E}_{\pi,P}[f(X_n)f(X_i)] \\
&= N\mathbf{E}_{\pi}(f^2) + 2\sum_{n=1}^{N}\sum_{i=1}^{k-1}\mathbf{E}_{\pi,P}[f(X_n)f(X_i)].
\end{aligned}$$

We can rewrite this second sum, using $n = i + k$, as

$$2\sum_{k=1}^{N}\sum_{i=1}^{N-k}\mathbf{E}_{\pi,P}[f(X_{i+k})f(X_i)].$$

As π is a stationary distribution and P is assumed to be time-homogeneous, this is equivalent to

$$2\sum_{k=1}^{N}(N-k)\mathbf{E}_{\pi,P}[f(X_k)f(X_0)].$$

As $\mathbf{E}_\pi(f^2) = \int_{x\in\mathbf{X}} f(x)^2\pi(dx) = \langle f, f\rangle$, and

$$\begin{aligned}\mathbf{E}_{\pi,P}(f(X_k)f(X_0)) &= \int_{x\in\mathbf{X}}\int_{y\in\mathbf{X}} f(x)f(y)\pi(dx)P^k(x,dy)\\ &= \int_{x\in\mathbf{X}} f(x)(\mathcal{P}^k f)(x)\pi(dx)\\ &= \langle f, \mathcal{P}^k f\rangle,\end{aligned}$$

we have proved the result. □

Theorem 29. *If P is a φ-irreducible reversible Markov kernel with stationary distribution π, then for every $f \in L_0^2(\pi)$,*

$$v(f,P) = \int_{\lambda\in\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda).$$

Note however, that this may still diverge to ∞.

Proof. Let $f \in L_0^2(\pi)$. Then by lemma 27, $E_{\mathcal{P}}(\{1\}) = 0$, so by lemma 28 and the spectral Theorem,

$$\begin{aligned}v(f,P) &= \lim_{N\to\infty}\left[\frac{1}{N}\mathbf{Var}\left(\sum_{n=1}^{N} f(X_n)\right)\right]\\ &= \lim_{N\to\infty}\left[\|f\|^2 + 2\sum_{k=1}^{N}\left(\frac{N-k}{N}\right)\langle f, \mathcal{P}^k f\rangle\right]\\ &= \|f\|^2 + 2\lim_{N\to\infty}\left[\int_{\lambda\in\sigma(\mathcal{P})}\sum_{k=1}^{N}\left(\frac{N-k}{N}\right)\lambda^k E_{\mathcal{P}}(d\lambda)\right]\\ &= \|f\|^2 + 2\lim_{N\to\infty}\left[\int_{\lambda\in[-1,1)}\sum_{k=1}^{\infty}\mathbf{1}_{k\le N}(k)\left(\frac{N-k}{N}\right)\lambda^k E_{\mathcal{P}}(d\lambda)\right].\end{aligned} \tag{1}$$

Notice that as for every $N \in \mathbf{N}$ and for every fixed $\lambda \in (-1,1)$,

$$\sum_{k=1}^{\infty}\left|\mathbf{1}_{k\le N}(k)\left(\frac{N-k}{N}\right)\lambda^k\right| \le \sum_{k=1}^{\infty}\left|\lambda^k\right| = \frac{|\lambda|}{1-|\lambda|} < \infty,$$

$$\begin{aligned}\lim_{N\to\infty}\sum_{k=1}^{\infty}\mathbf{1}_{k\le N}(k)\left(\frac{N-k}{N}\right)\lambda^k &= \sum_{k=1}^{\infty}\lim_{N\to\infty}\mathbf{1}_{k\le N}(k)\left(\frac{N-k}{N}\right)\lambda^k\\ &= \sum_{k=1}^{\infty}\lambda^k\\ &= \frac{\lambda}{1-\lambda}.\end{aligned} \tag{2}$$

So, letting $h_N^+ : (0,1) \to \mathbf{R}$ such that $h_N^+(\lambda) = \sum_{k=1}^\infty \mathbf{1}_{k\leq N}(k)\left(\frac{N-k}{N}\right)\lambda^k$ for every $\lambda \in (0,1)$, and letting $h^+ : (0,1) \to \mathbf{R}$ such that $h^+(\lambda) = \frac{\lambda}{1-\lambda}$ for every $\lambda \in (0,1)$, by equation 2, $h_N^+ \to h^+$ as $N \to \infty$. Furthermore, $h_N^+ \leq h_{N+1}^+$ for every $\lambda \in (0,1)$, for every $N \in \mathbf{N}$, thus as h_N^+ and h^+ are continuous and thus Borel measurable for every $N \in \mathbf{N}$, by the Monotone Convergence Theorem ([2] Theorem 2.14),

$$\lim_{N\to\infty}\left[\int_{\lambda\in(0,1)}\sum_{k=1}^\infty \mathbf{1}_{k\leq N}(k)\left(\frac{N-k}{N}\right)\lambda^k E_{\mathcal{P}}(d\lambda)\right] = \int_{\lambda\in(0,1)}\frac{\lambda}{1-\lambda}E_{\mathcal{P}}(d\lambda). \quad (3)$$

Simlarly, letting $h_N^- : (-1,0] \to \mathbf{R}$ such that $h_N^-(\lambda) = \sum_{k=1}^\infty \mathbf{1}_{k\leq N}(k)\left(\frac{N-k}{N}\right)\lambda^k$ for every $\lambda \in (-1,0]$, and letting $h^- : (-1,0] \to \mathbf{R}$ such that $h^-(\lambda) = \frac{\lambda}{1-\lambda}$ for every $\lambda \in (-1,0]$, by equation 2, $h_N^- \to h^-$ as $N \to \infty$. Furthermore, as $|h_N^-| \leq 2$ for every $N \in \mathbf{N}$, thus as h_N^- and h^- are continuous and thus Borel measurable for every $N \in \mathbf{N}$, by the Dominated Convergence Theorem ([2] Theorem 2.24),

$$\lim_{N\to\infty}\left[\int_{\lambda\in(-1,0]}\sum_{k=1}^\infty \mathbf{1}_{k\leq N}(k)\left(\frac{N-k}{N}\right)\lambda^k E_{\mathcal{P}}(d\lambda)\right] = \int_{\lambda\in(-1,0]}\frac{\lambda}{1-\lambda}E_{\mathcal{P}}(d\lambda). \quad (4)$$

As seen in [14], notice that for every $N \in \mathbf{N}$,

$$\begin{aligned}
&\sum_{k=1}^\infty \mathbf{1}_{k\leq N}(k)\frac{N-k}{N}(-1)^k E_{\mathcal{P}}(\{-1\}) \\
&\qquad = \sum_{k=1}^N \frac{N-k}{N}(-1)^k E_{\mathcal{P}}(\{-1\}) \\
&\qquad = \left(\frac{E_{\mathcal{P}}(\{-1\})}{N}\right)\sum_{k=1}^N (N-k)(-1)^k \\
&\qquad = \left(\frac{E_{\mathcal{P}}(\{-1\})}{N}\right)\sum_{k=1}^N [\mathbf{1}_{even}(k)(N-k) + \mathbf{1}_{odd}(k)(-N+k)] \\
&\qquad = \left(\frac{E_{\mathcal{P}}(\{-1\})}{N}\right)\sum_{m=1}^{\lfloor N/2\rfloor} [(N-2m) + (-N+2m-1)] \\
&\qquad = \left(\frac{E_{\mathcal{P}}(\{-1\})}{N}\right)\sum_{m=1}^{\lfloor N/2\rfloor} -1 \\
&\qquad = \left(\frac{\lfloor -N/2\rfloor}{N}\right)E_{\mathcal{P}}(\{-1\}).
\end{aligned}$$

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

$$\lim_{N\to\infty} \sum_{k=1}^{\infty} \mathbf{1}_{k\leq N}(k) \left(\frac{N-k}{N}\right) E_{\mathcal{P}}(\{-1\}) = \left(\frac{-1}{2}\right) E_{\mathcal{P}}(\{-1\}). \tag{5}$$

By equations 4 and 5, we know that

$$\left|\lim_{N\to\infty} \int_{\lambda\in(-1,0]} \sum_{k=1}^{\infty} \mathbf{1}_{k\leq N}(k) \left(\frac{N-k}{N}\right) \lambda^k E_{\mathcal{P}}(d\lambda)\right| = \left|\int_{\lambda\in(-1,0]} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda)\right| < \infty$$

$$\left|\lim_{N\to\infty} \sum_{k=1}^{\infty} \mathbf{1}_{k\leq N}(k) \left(\frac{N-k}{N}\right) E_{\mathcal{P}}(\{-1\})\right| = \left|\left(\frac{-1}{2}\right) E_{\mathcal{P}}(\{-1\})\right| < \infty,$$

so comining equations 3, 4 and 5,

$$\lim_{N\to\infty} \left[\int_{\lambda\in[-1,1)} \sum_{k=1}^{\infty} \mathbf{1}_{k\leq N}(k) \left(\frac{N-k}{N}\right) \lambda^k E_{\mathcal{P}}(d\lambda)\right] = \int_{\lambda\in[-1,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda).$$

Plugging this into 1, and as $E_{\mathcal{P}}(\{1\}) = 0$ by lemma 27, we have

$$\begin{aligned}
v(f,P) &= \|f\|^2 + 2 \lim_{N\to\infty} \left[\int_{\lambda\in[-1,1)} \sum_{k=1}^{\infty} \mathbf{1}_{k\leq N}(k) \left(\frac{N-k}{N}\right) \lambda^k E_{\mathcal{P}}(d\lambda)\right] \\
&= \|f\|^2 + 2 \int_{\lambda\in[-1,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \int_{\lambda\in\sigma(\mathcal{P})} E_{\mathcal{P}}(d\lambda) + 2 \int_{\lambda\in\sigma(\mathcal{P})} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \int_{\lambda\in\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda).
\end{aligned}$$

□

Notice that this integral may very well diverge to infinity. From the above proof, we notice that it is in deriving equation 3 that this integral may diverge.

Note that during the above proof, we also prove another useful characterization of the asymptotic variance for a reversible Markov chain. We note it here as a Proposition.

Definition (Periodicity). *A Markov kernel P with stationary distribution π is periodic if there exists $d \in \mathbf{N}$, $d \geq 2$ and $\{\mathcal{X}_k\}_{k=1}^{d} \subset \mathcal{F}$ such that $P(x, \mathcal{X}_{k+1}) = 1$ for π-a.e. $x \in \mathcal{X}_k$ for $k \in \{1, \ldots, d-1\}$ and $P(x, \mathcal{X}_1) = 1$ for π-a.e. $x \in \mathcal{X}_d$. If P is periodic, it's period is the largest such d such that the above holds.*

If P is not periodic, then P is called aperiodic.

Lemma 30. *If P is a reversible aperiodic Markov kernel reversible with respect to π, then -1 is not an eigenvalue of $\mathcal{P} : L^2_0(\pi) \to L^2_0(\pi)$.*

Proof. Assume for a contradiction that -1 is an eigenvalue of $\mathcal{P}$. Let $f \in L^2_0(\pi)$ be an eigenvector of $\mathcal{P}$ with eigenvalue -1 such that $\|f\| = 1$. By lemma 22 we can further assume that f is real valued. Take $d = 2$ and let $\mathcal{X}_1 = \{x \in \mathbf{X} : f(x) > 0\} = f^{-1}((0, \infty))$ and $\mathcal{X}_2 = \{x \in \mathbf{X} : f(x) < 0\} = f^{-1}((-\infty, 0))$. As f is $(\mathcal{F}, \mathcal{B}(\mathbf{R}))$-measureable, $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{F}$.

Notice that $\pi(\mathcal{X}_1), \pi(\mathcal{X}_2) > 0$ by definition of $\mathcal{X}_1$ and $\mathcal{X}_2$ as $f \neq 0$, as f is an eigenvector, and as

$$0 = \mathbf{E}_\pi(f) = \int_{\mathbf{X}} f(x)\pi(dx) = \int_{\mathcal{X}_1} f(x)\pi(dx) + \int_{\mathcal{X}_2} f(x)\pi(dx),$$

as $f \in L^2_0(\pi)$.

So, as $\mathcal{P}f = -f$, for π-a.e. $x \in \mathbf{X}$, $\int_{\mathbf{X}} f(y)P(x, dy) = -f(x)$, so

$$\int_{\mathcal{X}_1} f(y)P(x, dy) = -f(x) - \int_{\mathcal{X}_2} f(y)P(x, dy).$$

Thus as P is reversible with respect to π and $\int_{\mathcal{X}_1} f(x)\pi(dx) = -\int_{\mathcal{X}_2} f(x)\pi(dx)$,

$$\begin{aligned}
\int_{x\in\mathcal{X}_2}\int_{y\in\mathcal{X}_1} f(y)P(x,dy)\pi(dx) &= \int_{x\in\mathcal{X}_2}\left[-f(x) - \int_{y\in\mathcal{X}_2} f(y)P(x,dy)\right]\pi(dx) \\
\int_{x\in\mathcal{X}_2}\int_{y\in\mathcal{X}_1} f(y)P(y,dx)\pi(dy) &= -\int_{x\in\mathcal{X}_2} f(x)\pi(dx) - \iint_{x,y\in\mathcal{X}_2} f(y)P(y,dx)\pi(dy) \\
\int_{y\in\mathcal{X}_1} f(y)P(y,\mathcal{X}_2)\pi(dy) &= -\int_{x\in\mathcal{X}_2} f(x)\pi(dx) - \int_{y\in\mathcal{X}_2} f(y)P(y,\mathcal{X}_2)\pi(dy) \\
\int_{x\in\mathcal{X}_1} f(x)P(x,\mathcal{X}_2)\pi(dx) &= -\int_{x\in\mathcal{X}_2} f(x)\pi(dx) - \int_{x\in\mathcal{X}_2} f(x)P(x,\mathcal{X}_2)\pi(dx) \\
\int_{x\in\mathbf{X}} f(x)P(x,\mathcal{X}_2)\pi(dx) &= -\int_{x\in\mathcal{X}_2} f(x)\pi(dx) \\
\int_{x\in\mathbf{X}} f(x)P(x,\mathcal{X}_2)\pi(dx) &= \int_{x\in\mathcal{X}_1} f(x)\pi(dx).
\end{aligned}$$

Similarly, $\int_{\mathbf{X}} f(x)P(x, \mathcal{X}_1)\pi(dx) = \int_{\mathcal{X}_2} f(x)\pi(dx)$.

So assume for a contradiction that there exists $E \in \mathcal{F}$ such that $\pi(E) > 0$,

$E \subset \mathcal{X}_1$ and for every $x \in E$, $P(x, \mathcal{X}_2) < 1$. Then by definition of $\mathcal{X}_2$,

$$\begin{aligned}\int_{\mathbf{X}} f(x)P(x,\mathcal{X}_2)\pi(dx) &= \int_{\mathcal{X}_1} f(x)\pi(dx) \\ &> \int_{\mathcal{X}_1} f(x)P(x,\mathcal{X}_2)\pi(dx) \\ &\geq \int_{\mathbf{X}} f(x)P(x,\mathcal{X}_2)\pi(dx).\end{aligned}$$

So for π-a.e. $x \in \mathcal{X}_1$, $P(x, \mathcal{X}_2) = 1$. Similarly, for π-a.e. $x \in \mathcal{X}_2$, $P(x, \mathcal{X}_1) = 1$.

So, P is periodic with period $d \geq 2$, which contradicts the assumption that P is aperiodic. □

Proposition 31. *If P is an aperiodic φ-irreducible reversible Markov kernel, reversible with respect to π, then for every $f \in L_0^2(\pi)$,*

$$v(f,P) = \|f\|^2 + 2\sum_{k=1}^{\infty}\langle f, \mathcal{P}^k f\rangle \left(= \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k\right).$$

Proof. Let $f \in L_0^2(\pi)$. Then by lemma 30, -1 is not an eigenvalue of $\mathcal{P}$, so by lemma 26, $E_{\mathcal{P}}(\{-1\}) = 0$. So by theorem 29,

$$\begin{aligned}v(f,P) &= \int_{\lambda\in\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\ &= \int_{\lambda\in\sigma(\mathcal{P})} E_{\mathcal{P}}(d\lambda) + 2\int_{\lambda\in\sigma(\mathcal{P})} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\ &= \|f\|^2 + 2\int_{\lambda\in(-1,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda). \qquad (1)\end{aligned}$$

Let $h^+ : [0,1) \to \mathbf{R}$ such that $h^+(\lambda) = \frac{\lambda}{1-\lambda}$ for every $\lambda \in [0,1)$, and let $h_N^+ : [0,1) \to \mathbf{R}$ such that $h_N^+(\lambda) = \sum_{k=1}^N \lambda^k$ for every $\lambda \in [0,1)$. Then for every $\lambda \in [0,1)$, $h_N^+(\lambda) \to h(\lambda)$, and as $h_N^+ \leq h_{N+1}^+$ for every $N \in \mathbf{N}$, as h_N^+ is Borel-measureable for every N (as it is continuous), by the Monotone Convergence Theorem ([2] Theorem 2.14),

$$\int_{\lambda\in[0,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) = \sum_{k=1}^{\infty}\int_{\lambda\in[0,1)} \lambda^k E_{\mathcal{P}}(d\lambda).$$

Let $h^- : (-1,0) \to \mathbf{R}$ such that $h^-(\lambda) = \frac{\lambda}{1-\lambda}$ for every $\lambda \in (-1,0)$, and let $h_N^- : (-1,0) \to \mathbf{R}$ such that $h_N^-(\lambda) = \sum_{k=1}^N \lambda^k$ for every $\lambda \in (-1,0)$. Then for every $\lambda \in (-1,0)$, $h_N^-(\lambda) \to h(\lambda)$, and as $\left|h_N^-\right| \leq 2$ for every $N \in \mathbf{N}$, as h_N^- is

Borel-measureable for every N (as it is continuous), by the Dominated Convergence Theorem ([2] Theorem 2.24),

$$\int_{\lambda\in(-1,0)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) = \sum_{k=1}^{\infty} \int_{\lambda\in(-1,0)} \lambda^k E_{\mathcal{P}}(d\lambda).$$

So equation 1 becomes

$$\begin{aligned} v(f,P) &= \|f\|^2 + 2\int_{\lambda\in(-1,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\ &= \|f\|^2 + 2\sum_{k=1}^{\infty} \int_{\lambda\in(-1,1)} \lambda^k E_{\mathcal{P}}(d\lambda) \\ &= \|f\|^2 + 2\sum_{k=1}^{\infty} \int_{\lambda\in\sigma(\mathcal{P})} \lambda^k E_{\mathcal{P}}(d\lambda) \\ &= \|f\|^2 + 2\sum_{k=1}^{\infty} \langle f, \mathcal{P}^k f\rangle. \end{aligned}$$

□

We define here the identity operator $\mathcal{I} : L_0^2(\pi) \to L_0^2(\pi)$ such that $\mathcal{I}f = f$ for every $f \in L_0^2(\pi)$.

Proposition 32. *For a φ-irreducible reversible Markov kernel P reversible with respect to π, for every $f \in L_0^2(\pi)$,*

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

Note however that in many cases $(\mathcal{I}-\mathcal{P})^{-1} \notin \mathfrak{B}(L_0^2(\pi))$, i.e. $(\mathcal{I}-\mathcal{P})^{-1}$ is NOT a bounded operator. This is true whenever $1 \in \sigma(\mathcal{P})$.

Proof. By theorem 29, we have

$$\begin{aligned} v(f,P) &= \int_{\lambda\in\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\ &= \int_{\lambda\in\sigma(\mathcal{P})} E_{\mathcal{P}}(d\lambda) + 2\int_{\lambda\in\sigma(\mathcal{P})} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\ &= \|f\|^2 + 2\langle f, \mathcal{P}(\mathcal{I}-\mathcal{P})^{-1} f\rangle, \end{aligned}$$

as $\mathcal{P}(\mathcal{I}-\mathcal{P})^{-1} = \int_{\lambda\in\sigma(\mathcal{P})} \frac{\lambda}{1-\lambda} \mathcal{E}_{\mathcal{P}}(d\lambda)$. □

Remark. *Note however that in many cases $(\mathcal{I}-\mathcal{P})^{-1} \notin \mathfrak{B}(L_0^2(\pi))$, i.e. $(\mathcal{I}-\mathcal{P})^{-1}$ is NOT a bounded operator. This is true whenever $1 \in \sigma(\mathcal{P})$.*

4 Inverse Inequalities, Lemma 6 of Neal Paper

Lemma 33. *If X, Y, and Z are bounded linear operators on a Hilbert space $\mathbf{H}$ such that $\langle f, Xf\rangle \leq \langle f, Yf\rangle$ for every $f \in \mathbf{H}$, and Z is self-adjoint, then $\langle f, ZXZf\rangle \leq \langle f, ZYZf\rangle$ for every $f \in \mathbf{H}$.*

Proof. For every $f \in \mathbf{H}$, $Zf \in \mathbf{H}$, so

$$\langle f, ZXZf\rangle = \langle Zf, XZf\rangle \leq \langle Zf, YZf\rangle = \langle f, ZYZf\rangle.$$

$\square$

In the finite state space case, $L_0^2(\pi)$ is a finite dimensional vector space, and thus in order to prove that $\langle f, Tf\rangle \leq \langle f, Nf\rangle$ for every $f \in \mathbf{V}$ if and only if $\langle f, T^{-1}f\rangle \geq \langle f, N^{-1}f\rangle$ for every $f \in \mathbf{V}$, when T and N are self-adjoint operators, the only additional assumption needed is that T and N are strictly positive (next definition). This is presented by Rosenthal and Neal in [1], section 8. However, in the general case, as $L_0^2(\pi)$ may not be finite dimensional, T and N being strictly positive is not enough. So, we must use a slightly stronger assumption, which in the finite case, is equivalent to being strictly positive.

Definition (Positive Operators)**.** *We say a bounded linear self-adjoint operator T on a Hilbert space $\mathbf{H}$ is (strictly) positive, denoted $T > 0$ ($T \geq 0$), if for every $f \in \mathbf{H}$, $f \neq 0$,*

$$\langle f, Tf\rangle > 0, \qquad (\langle f, Tf\rangle \geq 0)\,.$$

If T is strictly positive, or even just positive, then no negative numbers will be in it's spectrum.

Lemma 34. *If T is a positive bounded linear self-adjoint operator on a Hilbert space $\mathbf{H}$, then $\sigma(T) \subset [0, \infty)$.*

Proof. If $\lambda < 0$, then for every $f \in \mathbf{H}$ such that $f \neq 0$,

$$\begin{aligned}
|\langle (T-\lambda)f, f\rangle| &= |\langle Tf, f\rangle - \lambda \left\|f\right\|^2| \\
&\geq \langle Tf, f\rangle + |\lambda| \left\|f\right\|^2 && (\text{as } \lambda < 0 \text{ and } T \geq 0) \\
&\geq |\lambda| \left\|f\right\|^2. && (\text{as } T \geq 0)
\end{aligned}$$

Furthermore, by the Cauchy-Schwarz inequality,

$$|\langle (T-\lambda)f, f\rangle| \leq \left\|(T-\lambda)f\right\| \left\|f\right\|.$$

Thus, as $\lambda \neq 0$, we have

$$0 < |\lambda| \left\|f\right\| \leq \left\|(T-\lambda)f\right\|.$$

So, by lemma 16, $T - \lambda$ is invertible, and $\lambda \notin \sigma(T)$ by definition. $\square$

Lemma 35. *If* $\mathbf{V}$ *is a finite dimensional vector space and* $T \in \mathfrak{B}(\mathbf{V})$ *is self-adjoint and strictly positive, then* T *is invertible. I.e.* $0 \notin \sigma(T)$.

Proof. Let $f \in \mathbf{V}$ such that $f \neq 0$. Then by the Cauchy-Schwartz inequality, as T is strictly positive,

$$0 < \langle f, Tf \rangle \leq \|Tf\| \, \|f\| \, .$$

So, $\|Tf\| \neq 0$, and thus $Tf \neq 0$. So, $\ker(T) = \{0\}$, and thus T is injective. So as $\mathbf{V}$ is a finite vector space, by theorem 12, T is invertible. □

In general however, being strictly positive doesn't mean that T is invertible. Although the spectrum won't have any negative numbers, 0 may still exist in the spectrum even if it is strictly positive. This is why we must assume something stronger, which in the finite case is equivalent to being strictly positive.

What we assume is that $T \in \mathfrak{B}(\mathbf{H})$ is self-adjoint such that $\sigma(T) \subset (0, \infty)$. As $0 \notin \sigma(T)$, T is invertible, and as $T = \int_{\sigma(T)} \lambda \mathcal{E}(d\lambda)$ is the spectral decomposition of T, for every $f \in \mathbf{H}$,

$$\langle f, Tf \rangle = \int_{\sigma(T)} \lambda E_{T,f}(d\lambda) = \int_{(0,\infty)} \lambda E_{T,f}(d\lambda) > 0.$$

(NOTE: This lemma is also proved in section 5 using the open-mapping theorem, as that is the justification Rudin gave for it in [7]. We prove it more directly here.)

Lemma 36. *If* T *is a normal bounded linear operator on a Hilbert space* $\mathbf{H}$*, then there exists* $\delta > 0$ *such that* $\delta \|f\| \leq \|Tf\|$ *for every* $f \in \mathbf{H}$ *if and only if* T *is invertible.*

Remark. *The forward implication is lemma 16.*

Proof. We've already done the only if direction in lemma 16. I copy and paste it here:

First we notice that given the hypothesis, T is injective, as for every $f \in \mathbf{H}$ such that $f \neq 0$, $\|Tf\| \geq \delta \|f\| > 0$, so $Tf \neq 0$. Thus, by Proposition 14, range(T) is dense in $\mathbf{H}$. Next notice that by Proposition 15, range(T) is closed in $\mathbf{H}$. So,

$$\mathbf{H} = \overline{\operatorname{range}(T)} = \operatorname{range}(T).$$

So, T is invertible.

For the converse, say T is invertible. Then let $\delta = \|T^{-1}\|^{-1}$. Then for every $f \in \mathbf{H}$,

$$\|f\| = \left\|T^{-1}Tf\right\| \leq \left\|T^{-1}\right\| \|Tf\| \, ,$$

and the result follows. □

Lemma 37. *If T and N are self-adjoint bounded linear operators on a Hilbert space $\mathbf{H}$, such that $\sigma(T), \sigma(N) \subset (0, \infty)$, then $\langle f, Tf \rangle \leq \langle f, Nf \rangle$ for every $f \in \mathbf{H}$, if and only if $\langle f, T^{-1} f \rangle \geq \langle f, N^{-1} f \rangle$, for ever $f \in \mathbf{H}$.*

Proof. Say $\langle f, Tf \rangle \leq \langle f, Nf \rangle$ for every $f \in \mathbf{H}$.

As $0 \notin \sigma(N)$, N is invertible, and so

$$N^{-1/2} := \int_{\lambda \in \sigma(N^{-1})} \lambda^{1/2} \mathcal{E}_{N^{-1}}(d\lambda) = \int_{\lambda \in \sigma(N)} \lambda^{-1/2} \mathcal{E}_N(d\lambda)$$

is a well defined bounded self-adjoint linear operator. Similarly, $T^{1/2} = \int_{\lambda \in \sigma(T)} \lambda^{1/2} \mathcal{E}_T(d\lambda)$ is a well defined bounded self-adjoint linear operator.

So, for every $f \in \mathbf{H}$, we have

$$\langle f, N^{-1/2} T N^{-1/2} f \rangle = \langle T^{1/2} N^{-1/2} f, T^{1/2} N^{-1/2} f \rangle = \left\| T^{1/2} N^{-1/2} f \right\|^2 \geq 0.$$

Furthermore, as $\sigma(T) \subset (0, \infty)$, T is invertible, so by lemma 36 there exists $\delta_T > 0$ such that $\|Tf\| \geq \delta_T \|f\|$ for every $f \in \mathbf{H}$. Also, notice that $\sigma(N^{-1/2}) \subset (0, \infty)$, thus by lemma 36, there exists $\delta_1 > 0$ such that $\left\| N^{-1/2} f \right\| \geq \delta_1 \|f\|$ for every $f \in \mathbf{H}$. So, for every $f \in \mathbf{H}$,

$$\left\| N^{-1/2} T N^{-1/2} f \right\| \geq \delta_1 \left\| T N^{-1/2} f \right\| \geq \delta_1 \delta_T \left\| N^{-1/2} f \right\| \geq \delta_1 \delta_T \delta_1 \|f\|,$$

so by lemma 36, $N^{-1/2} T N^{-1/2}$ is invertible, and thus $0 \notin \sigma(N^{-1/2} T N^{-1/2})$.

By using lemma 33 with $X = T$, $Y = N$ and $Z = N^{-1/2}$, for every $f \in \mathbf{H}$,

$$\langle f, N^{-1/2} T N^{-1/2} f \rangle \leq \langle f, N^{-1/2} N N^{-1/2} f \rangle = \|f\|^2.$$

So if $\lambda > 1$, for any $f \in \mathbf{H}$, as $0 \leq \langle N^{-1/2} T N^{-1/2} f, f \rangle \leq \|f\|^2$, by the Cauchy-Schwartz inequality,

$$\begin{aligned} \left\| (N^{-1/2} T N^{-1/2} - \lambda) f \right\| \|f\| &\geq |\langle (N^{-1/2} T N^{-1/2} - \lambda) f, f \rangle| \\ &= |\langle N^{-1/2} T N^{-1/2} f, f \rangle - \lambda \|f\|^2 | \\ &\geq |1 - \lambda| \|f\|^2, \end{aligned}$$

so $\left\| (N^{-1/2} T N^{-1/2} - \lambda) f \right\| \geq |1 - \lambda| \|f\|$, and as $|1 - \lambda| > 0$, by lemma 36, $(N^{-1/2} T N^{-1/2} - \lambda)$ is invertible, so $\lambda \notin \sigma(N^{-1/2} T N^{-1/2})$.

Thus we have $\sigma(N^{-1/2} T N^{-1/2}) \subset (0, 1]$.

As $0 \notin \sigma(N^{-1/2} T N^{-1/2})$, $N^{-1/2} T N^{-1/2}$ is invertible. Let K denote the inverse. I.e. let $K = (N^{-1/2} T N^{-1/2})^{-1}$. Furthermore, we have $\sigma(K) \subset [1, \infty)$. So for every $f \in \mathbf{H}$, $\|f\|^2 \leq \langle f, Kf \rangle$.

So by using lemma 33, with $X = \mathcal{I}$, $Y = K$ and $Z = N^{-1/2}$, for every $f \in \mathbf{H}$,

$$\begin{aligned}
\langle f, N^{-1} f\rangle &= \langle f, N^{-1/2}\mathcal{I}N^{-1/2} f\rangle \\
&\leq \langle f, N^{-1/2}KN^{-1/2} f\rangle \\
&= \langle f, N^{-1/2}(N^{-1/2}TN^{-1/2})^{-1}N^{-1/2} f\rangle \\
&= \langle f, N^{-1/2}N^{1/2}T^{-1}N^{1/2}N^{-1/2} f\rangle \qquad \text{(same as finite case)} \\
&= \langle f, T^{-1} f\rangle.
\end{aligned}$$

For the other direction, replace N with T^{-1} and T with N^{-1}. □

5 *Understanding the Open Mapping Theorem

Recalling lemma 36, Rudin uses an important Theorem in functional analysis, the open mapping Theorem, to complete the only if direction. To understand the open mapping Theorem, we need the following results and definitions. There are competing definitions, one is given in Rudin's Functional Analysis, [7], the other is given in Munkres' Topology, [9]. I present both versions here for completeness, though they are equivalent. The first definition can be found in Rudin's book [7].

Definition (First and Second Category Sets). *Given a topological space $(X, \mathcal{T})$, a subset $A \subset X$ is said to be of the first category if it is the countable union of nowhere dense sets, sets whose closure has empty interior. Sets not of the first category are of the second category.*

A perhaps easier to understand, though (I think) lesser known equivalent definition is that of Baire spaces, as seen in Munkres' book, [9].

Definition (Baire Spaces). *A topological space $(X, \mathcal{T})$ is a Baire space if given any countable collection of dense open sets of X, $\{U_n\} \subset \mathcal{T}$ such that $\overline{U_n} = X$ for every $n \in \mathbf{N}$, their intersection is also dense in X, i.e.*

$$\overline{\cap U_n} = X.$$

Note that Baire spaces are of the second category. Thus,

Theorem 38 (Baire's Theorem). *If X is a complete metric space, then X is a Baire space, or X is of the second category.*

Proof. Half of Theorem 48.2 of [9]. □

Now we can understand the open-mapping theorem. Though it is much more general than this, I present a version more direct for our purpose.

Definition (Open Map). *A function $f : X \to Y$, where X and Y are topological spaces, is an open map if for every open set $U \subset X$, $f(U) \subset Y$ is open.*

Theorem 39 (Open Mapping Theorem). *If $T : \mathbf{H} \to \mathbf{H}$ is a bounded linear operator on a Hilber space $\mathbf{H}$, then T is an open mapping if $T(\mathbf{H})$ is a Baire space.*

Proof. A simplified version of Theorem 2.11 of [7]. □

Now we are able to prove the converse of lemma 16, the only if direction of lemma 36, with the open mapping Theorem.

Lemma 40 (Lemma 36 again). *If T is a normal bounded linear operator on a Hilbert space $\mathbf{H}$, then there exists $\delta > 0$ such that $\delta \|f\| \leq \|Tf\|$ for $f \in \mathbf{H}$ if and only if T is invertible.*

Proof. We've already done the only if direction in lemma 16. I copy and paste it here: First we notice that given the hypothesis, T is injective, as for every $f \in \mathbf{H}$ such that $f \neq 0$, $\|Tf\| \geq \delta \|f\| > 0$, so $Tf \neq 0$. Thus, by Proposition 14, range(T) is dense in $\mathbf{H}$. Next notice that by Proposition 15, range(T) is closed in $\mathbf{H}$. So,

$$\mathbf{H} = \overline{\text{range}(T)} = \text{range}(T).$$

So, T is invertible.

Say T is invertible. Then as T is bounded, it is continuous by lemma 5. Furthermore, by definition of a Hilbert space, $\mathbf{H}$ is a complete metric space. Thus, by Baire's theorem, theorem 38, $\mathbf{H}$ is a Baire space.

So, as T is invertible, it is surjective. Thus, $T(\mathbf{H}) = \mathbf{H}$, and $T(\mathbf{H})$ is a Baire space. So by the open mapping theorem, theorem 39, T is an open map.

Thus T^{-1} is a continuous map, and thus is bounded by lemma 5, so for any $f \in \mathbf{H}$,

$$\|f\| = \left\|T^{-1}Tf\right\| \leq \left\|T^{-1}\right\| \|Tf\| .$$

□

6 Efficiency Dominance

Definition (Efficiency Dominance). *Given two Markov kernels P and Q on $(\mathbf{X}, \mathcal{F})$ with stationary distribution π, we say that P efficiency-dominates Q if for every $f \in L^2(\pi)$,*

$$v(f, P) \leq v(f, Q).$$

Theorem 41. *If P and Q are φ-irreducible reversible Markov kernels, reversible with respect to π, then P efficiency dominates Q if and only if for every $f \in L_0^2(\pi)$,*

$$\langle f, \mathcal{P} f\rangle \leq \langle f, \mathcal{Q} f\rangle.$$

Proof. Say $\langle f, \mathcal{P} f\rangle \leq \langle f, \mathcal{Q} f\rangle$ for every $f \in L_0^2(\pi)$. For every $\eta \in [0,1)$, let $T_{\mathcal{P},\eta} = \mathcal{I} - \eta\mathcal{P}$ and $T_{\mathcal{Q},\eta} = \mathcal{I} - \eta\mathcal{Q}$. Then as $\|\mathcal{P}\|, \|\mathcal{Q}\| \leq 1$, for every $f \in L_0^2(\pi)$, by the Cauchy-Schwartz inequality, $|\langle f, \mathcal{P} f\rangle| \leq \|\mathcal{P} f\| \|f\| \leq \|\mathcal{P}\| \|f\|^2 \leq \|f\|^2$, and similarly $|\langle f, \mathcal{Q} f\rangle| \leq \|f\|^2$, so again by the Cauchy-Schwartz inequality,

$$\begin{aligned}\|T_{\mathcal{P},\eta} f\| \|f\| &\geq |\langle T_{\mathcal{P},\eta} f, f\rangle| \\ &= \left|\|f\|^2 - \eta\langle f, \mathcal{P} f\rangle\right| \\ &\geq |1-\eta| \|f\|^2,\end{aligned}$$

so $\|T_{\mathcal{P},\eta} f\| \geq |1-\eta| \|f\|$ for every $f \in L_0^2(\pi)$, and similarly $\|T_{\mathcal{Q},\eta} f\| \geq |1-\eta| \|f\|$ for every $f \in L_0^2(\pi)$. As $\eta \in [0,1)$, $|1-\eta| > 0$, so by lemma 16, $T_{\mathcal{P},\eta}$ and $T_{\mathcal{Q},\eta}$ are both invertible, so $0 \notin \sigma(T_{\mathcal{P},\eta}), \sigma(T_{\mathcal{Q},\eta})$.

So, as $\sigma(\mathcal{P}), \sigma(\mathcal{Q}) \subset [-1,1]$, $\sigma(T_{\mathcal{P},\eta}), \sigma(T_{\mathcal{Q},\eta}) \subset (0,2) \subset (0,\infty)$ for every $\eta \in [0,1)$.

So for every $\eta \in [0,1)$, as $T_{\mathcal{P},\eta}$ and $T_{\mathcal{Q},\eta}$ are both self-adjoint, and for every $f \in L_0^2(\pi)$,

$$\langle f, T_{\mathcal{Q},\eta} f\rangle = \|f\|^2 - \eta\langle f, \mathcal{Q} f\rangle \leq \|f\|^2 - \eta\langle f, \mathcal{P} f\rangle = \langle f, T_{\mathcal{P},\eta} f\rangle,$$

by lemma 37, for every $f \in L_0^2(\pi)$, $\langle f, T_{\mathcal{Q},\eta}^{-1} f\rangle \geq \langle f, T_{\mathcal{P},\eta}^{-1} f\rangle$.

So, for every $f \in L_0^2(\pi)$,

$$\begin{aligned}\|f\|^2 + \eta\langle f, \mathcal{P}(\mathcal{I} - \eta\mathcal{P})^{-1} f\rangle &= \langle f, T_{\mathcal{P},\eta}^{-1} f\rangle \\ &\leq \langle f, T_{\mathcal{Q},\eta}^{-1} f\rangle \\ &= \|f\|^2 + \eta\langle f, \mathcal{Q}(\mathcal{I} - \eta\mathcal{Q})^{-1} f\rangle,\end{aligned}$$

so for every $f \in L_0^2(\pi)$, $\langle f, \mathcal{P}(\mathcal{I} - \eta\mathcal{P})^{-1} f\rangle \leq \langle f, \mathcal{Q}(\mathcal{I} - \eta\mathcal{Q})^{-1} f\rangle$.

Let $f \in L_0^2(\pi)$. For every $\eta \in [0,1)$ let $h_\eta^+ : [0,1) \to \mathbf{R}$ such that $h_\eta^+(\lambda) = \frac{\lambda}{1-\eta\lambda}$ for every $\lambda \in [0,1)$. Then let $h^+ : [0,1) \to \mathbf{R}$ such that $h^+(\lambda) = \frac{\lambda}{1-\lambda}$ for every $\lambda \in [0,1)$. Then as $h_{\eta_1}^+ \leq h_{\eta_2}^+$ whenever $\eta_1 \leq \eta_2$, and $h_\eta^+ \to h^+$ as $\eta \to 1^-$, by the Monotone Convergence Theorem ([2] Theorem 2.14),

$$\lim_{\eta\to 1^-} \int_{[0,1)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{P}}(d\lambda) = \int_{[0,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda)$$

and

$$\lim_{\eta\to 1^-} \int_{[0,1)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{Q}}(d\lambda) = \int_{[0,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{Q}}(d\lambda).$$

Similarly, for every $\eta \in [0,1)$ let $h_\eta^- : [-1,0) \to \mathbf{R}$ such that $h_\eta^-(\lambda) = \frac{\lambda}{1-\eta\lambda}$ for every $\lambda \in [-1,0)$. Then let $h^- : [-1,0) \to \mathbf{R}$ such that $h^-(\lambda) = \frac{\lambda}{1-\lambda}$ for every $\lambda \in [-1,0)$. Then as $|h_\eta^-| \leq 1$ for every $\eta \in [0,1)$, and $h_\eta^- \to h^-$ as $\eta \to 1^-$, by the Dominated Convergence Theorem ([2] Theorem 2.24),

$$\lim_{\eta \to 1^-} \int_{[-1,0)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{P}}(d\lambda) = \int_{[-1,0)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda)$$

and

$$\lim_{\eta \to 1^-} \int_{[-1,0)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{Q}}(d\lambda) = \int_{[-1,0)} \frac{\lambda}{1-\lambda} E_{\mathcal{Q}}(d\lambda).$$

So, as P is φ-irreducible, by proposition 27, $E_{\mathcal{P}}(\{1\}) = 0$. So, by theorem 29,

$$\begin{aligned}
v(f,P) &= \int_{\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2\int_{\sigma(\mathcal{P})} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2\int_{[-1,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2\int_{[-1,0)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) + 2\int_{[0,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2\lim_{\eta \to 1^-}\int_{[-1,0)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{P}}(d\lambda) + 2\lim_{\eta \to 1^-}\int_{[0,1)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2\lim_{\eta \to 1^-}\int_{[-1,1)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2\lim_{\eta \to 1^-} \langle f, \mathcal{P}\left(\mathcal{I} - \eta\mathcal{P}\right)^{-1} f\rangle \\
&\leq \|f\|^2 + 2\lim_{\eta \to 1^-} \langle f, \mathcal{Q}\left(\mathcal{I} - \eta\mathcal{Q}\right)^{-1} f\rangle \\
&= \|f\|^2 + 2\lim_{\eta \to 1^-}\int_{[-1,1)} \frac{\lambda}{1-\eta\lambda} E_{\mathcal{Q}}(d\lambda) \\
&= \|f\|^2 + 2\int_{[-1,1)} \frac{\lambda}{1-\lambda} E_{\mathcal{Q}}(d\lambda) \\
&= \int_{\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{\mathcal{P}}(d\lambda) \\
&= v(f,Q).
\end{aligned}$$

For the converse, just as above, for every $f \in L_0^2(\pi)$,

$$\begin{aligned}
\|f\|^2 + 2 \lim_{\eta \to 1^-} \langle f, \mathcal{P} \left(\mathcal{I} - \eta \mathcal{P}\right)^{-1} f \rangle &= v(f, P) \\
&\leq v(f, Q) \\
&= \|f\|^2 + 2 \lim_{\eta \to 1^-} \langle f, \mathcal{Q} \left(\mathcal{I} - \eta \mathcal{Q}\right)^{-1} f \rangle.
\end{aligned}$$

So, for every $f \in L_0^2(\pi)$,

$$\begin{aligned}
\lim_{\eta \to 1^-} \langle f, T_{\mathcal{P},\eta}^{-1} f \rangle &= \|f\|^2 + \lim_{\eta \to 1^-} \eta \langle f, \mathcal{P} \left(\mathcal{I} - \eta \mathcal{P}\right)^{-1} f \rangle \\
&\leq \|f\|^2 + \lim_{\eta \to 1^-} \eta \langle f, \mathcal{Q} \left(\mathcal{I} - \eta \mathcal{Q}\right)^{-1} f \rangle \\
&= \lim_{\eta \to 1^-} \langle f, T_{\mathcal{Q},\eta}^{-1} f \rangle.
\end{aligned}$$

So by lemma 37, for every $f \in L_0^2(\pi)$,

$$\langle f, \left(\mathcal{I} - \mathcal{P}\right) f \rangle = \lim_{\eta \to 1^-} \langle f, T_{\mathcal{P},\eta} f \rangle \geq \lim_{\eta \to 1^-} \langle f, T_{\mathcal{Q},\eta} f \rangle = \langle f, \left(\mathcal{I} - \mathcal{Q}\right) f \rangle,$$

and thus for every $f \in L_0^2(\pi)$,

$$\langle f, \mathcal{P} f \rangle \leq \langle f, \mathcal{Q} f \rangle.$$

Furthermore, for any $f \in L^2(\pi)$, $f_0 := f - \mathbf{E}_\pi(f) \in L_0^2(\pi)$ by the linearity of the expected value functional, and notice that

$$\begin{aligned}
v(f, P) &= \lim_{N \to \infty} \left[\frac{1}{N} \mathbf{Var} \left(\sum_{n=1}^{N} f(X_n) \right) \right] \\
&= \lim_{N \to \infty} \left[\frac{1}{N} \mathbf{Var} \left(\sum_{n=1}^{N} f_0(X_n) + N \mathbf{E}_\pi(f) \right) \right] \\
&= \lim_{N \to \infty} \left[\frac{1}{N} \mathbf{Var} \left(\sum_{n=1}^{N} f_0(X_n) \right) \right] \\
&= v(f_0, P),
\end{aligned}$$

and similarly $v(f, Q) = v(f_0, Q)$. Thus $v(f_0, P) \leq v(f_0, Q)$ for every $f_0 \in L_0^2(\pi)$ if and only if $v(f, P) \leq v(f, Q)$ for every $f \in L^2(\pi)$. $\square$

Lemma 42. *If T is a bounded self-adjoint linear operator on a Hilbert space $\mathbf{H}$, then $\langle f, Tf \rangle \geq 0$ for every $f \in \mathbf{H}$ if and only if $\sigma(T) \subset [0, \infty)$.*

Proof. The only if direction is the same as lemma 34:

If $\lambda < 0$, then for every $f \in \mathbf{H}$ such that $f \neq 0$, by the Cauchy-Schwartz inequality,

$$\begin{aligned}\|(T-\lambda)f\|\,\|f\| &\geq |\langle (T-\lambda)f, f\rangle| \\ &= |\langle Tf, f\rangle - \lambda \|f\|^2| \\ &\geq \langle Tf, f\rangle + |\lambda|\,\|f\|^2 && \text{(as } \lambda < 0 \text{ and } T \geq 0) \\ &\geq |\lambda|\,\|f\|^2. && \text{(as } T \geq 0)\end{aligned}$$

Thus as $f \neq 0$ and $\lambda \neq 0$,

$$\|(T-\lambda)f\| \geq |\lambda|\,\|f\| > 0.$$

So, by lemma 16, $T-\lambda$ is invertible, and thus $\lambda \notin \sigma(T)$ by definition.

If $\sigma(T) \subset [0,\infty)$, then for every $f \in \mathbf{H}$,

$$\begin{aligned}\langle f, Tf\rangle &= \int_{\lambda\in\sigma(T)} \lambda E_T(d\lambda) \\ &= \int_{\lambda\in[0,\infty)} \lambda E_T(d\lambda) \\ &\geq \int_{\lambda\in[0,\infty)} E_T(d\lambda) \\ &\geq 0,\end{aligned}$$

as E_T is a positive measure by lemma 20. $\square$

Theorem 43. *If P and Q are φ-irreducible reversible Markov kernels reversible with respect to π, then P efficiency dominates Q if and only if $\sigma(\mathcal{Q}-\mathcal{P}) \subset [0,\infty)$.*

Proof. By theorem 41, P efficiency-dominates Q if and only if $\langle f, \mathcal{P}f\rangle \leq \langle f, \mathcal{Q}f\rangle$ for every $f \in L_0^2(\pi)$. This is equivalent to

$$\langle f, (\mathcal{Q}-\mathcal{P})f\rangle \geq 0, \quad \text{for every } f \in L_0^2(\pi),$$

as $-\infty < \langle f, \mathcal{P}f\rangle, \langle f, \mathcal{Q}f\rangle < \infty$ as $\mathcal{P}, \mathcal{Q} \in \mathfrak{B}(L_0^2(\pi))$.

Thus as $\mathcal{Q}-\mathcal{P}$ is a bounded self-adjoint linear operator on $L_0^2(\pi)$, by lemma 42, $\langle f, (\mathcal{Q}-\mathcal{P})f\rangle \geq 0$ for every $f \in L_0^2(\pi)$ if and only if $\sigma(\mathcal{Q}-\mathcal{P}) \subset [0,\infty)$. $\square$

Theorem 44. *Efficiency dominance is a partial order on φ-irreducible reversible Markov kernels with the same reversible distribution. (As shown in [5].)*

Proof. Reflexivity is trivial.

Suppose P and Q are φ-irreducible reversible Markov kernels reversible with respect to π such that P efficiency dominates Q and Q efficiency dominates P. Then by theorem 41, for every $f \in L_0^2(\pi)$,

$$\langle f, \mathcal{P}f \rangle \leq \langle f, \mathcal{Q}f \rangle \quad \text{and} \quad \langle f, \mathcal{P}f \rangle \geq \langle f, \mathcal{Q}f \rangle,$$

so $\langle f, \mathcal{P}f \rangle = \langle f, \mathcal{Q}f \rangle$ for every $f \in L_0^2(\pi)$. Thus $\langle f, (\mathcal{Q} - \mathcal{P})f \rangle = 0$ for every $f \in L_0^2(\pi)$. So for every $g, h \in L_0^2(\pi)$, as $\mathcal{Q}$ and $\mathcal{P}$ are self-adjoint by theorem 9,

$$\begin{aligned} 0 &= \langle g + h, (\mathcal{Q} - \mathcal{P})(g + h) \rangle \\ &= \langle g, (\mathcal{Q} - \mathcal{P})g \rangle + \langle h, (\mathcal{Q} - \mathcal{P})h \rangle + 2\langle g, (\mathcal{Q} - \mathcal{P})h \rangle \\ &= 2\langle g, (\mathcal{Q} - \mathcal{P})h \rangle. \end{aligned}$$

So for every $g, h \in L_0^2(\pi)$, $\langle g, (\mathcal{Q} - \mathcal{P})h \rangle = 0$. Thus $\mathcal{Q} - \mathcal{P} = 0$, so $\mathcal{P} = \mathcal{Q}$, and thus $P = Q$. So the relation is antisymmetric.

Suppose P, Q and R are φ-irreducible reversible Markov kernels reversible with respect to π, such that P efficiency dominates Q and Q efficiency dominates R. Then by theorem 43, $\sigma(\mathcal{Q} - \mathcal{P}), \sigma(\mathcal{R} - \mathcal{Q}) \subset [0, \infty)$, so by lemma 42, for every $f \in L_0^2(\pi)$, $\langle f, (\mathcal{R} - \mathcal{Q})f \rangle \geq 0$ and $\langle f, (\mathcal{Q} - \mathcal{P})f \rangle \geq 0$. So, for every $f \in L_0^2(\pi)$,

$$\langle f, (\mathcal{R} - \mathcal{P})f \rangle = \langle f, (\mathcal{R} - \mathcal{Q})f \rangle + \langle f, (\mathcal{Q} - \mathcal{P})f \rangle \geq 0.$$

And thus by lemma 42 followed by theorem 43, we have P efficiency dominates R, so the relation is transitive. □

7 Mixing Kernels

Lemma 45. *If P is a φ-irreducible Markov kernel, then for every Markov kernel R and for every $\alpha \in (0, 1]$, $\alpha P + (1 - \alpha)R$ is an irreducible Markov kernel.*

Proof. Obviously $\alpha P + (1 - \alpha)R$ is a Markov kernel. As P is φ-irreducible, there exists a σ-finite measure φ on $(\mathbf{X}, \mathcal{F})$ such that the definition of φ-irreducibility holds for P. So, for every $E \in \mathcal{F}$ such that $\varphi(E) > 0$, for every $x \in \mathbf{X}$, there exists $n \in \mathbf{N}$ such that

$$(\alpha P + (1 - \alpha)R)^n (x, E) \geq \alpha^n P^n(x, E) > 0.$$

So $\alpha P + (1 - \alpha)R$ satisfies the definition of φ-irreducibility with the same σ-finite measure φ. □

Theorem 46. *Let P and Q be φ-irreducible reversible Markov kernels, reversible with respect to π, and let R be a reversible Markov kernel, also reversible with respect to π. Then for every $\alpha \in (0,1)$, P efficiency dominates Q if and only if $\alpha P + (1-\alpha)R$ efficiency dominates $\alpha Q + (1-\alpha)R$.*

Proof. By theorem 41, P efficiency dominates Q if and only if for every $f \in L_0^2(\pi)$,

$$\langle f, \mathcal{P} f\rangle \leq \langle f, \mathcal{Q} f\rangle.$$

Notice that for every $f \in L_0^2(\pi)$,

$$\langle f, (\alpha\mathcal{P} + (1-\alpha)\mathcal{R})\, f\rangle = \alpha\langle f, \mathcal{P} f\rangle + (1-\alpha)\langle f, \mathcal{R} f\rangle,$$

and

$$\langle f, (\alpha\mathcal{Q} + (1-\alpha)\mathcal{R})\, f\rangle = \alpha\langle f, \mathcal{Q} f\rangle + (1-\alpha)\langle f, \mathcal{R} f\rangle,$$

so this is equivalent to

$$\langle f, (\alpha\mathcal{P} + (1-\alpha)\mathcal{R})\, f\rangle \leq \langle f, (\alpha\mathcal{Q} + (1-\alpha)\mathcal{R})\, f\rangle$$

for every $f \in L_0^2(\pi)$.

By lemma 45, $\alpha P + (1-\alpha)R$ and $\alpha Q + (1-\alpha)R$ are also φ-irreducible. Thus as they are also reversible with respect to π, by theorem 41, $\alpha P + (1-\alpha)R$ efficiency dominates $\alpha Q + (1-\alpha)R$ if and only if $\langle f, (\alpha\mathcal{P} + (1-\alpha)\mathcal{R})\, f\rangle \leq \langle f, (\alpha\mathcal{Q} + (1-\alpha)\mathcal{R})\, f\rangle$ for every $f \in L_0^2(\pi)$. □

Theorem 47. *Let $P_1, \ldots, P_l$ and $Q_1, \ldots, Q_l$ be Markov kernels reversible with respect to π. Let $\alpha_1, \ldots, \alpha_l$ be mixing probabilities, i.e. $\alpha_k \geq 0$ for every k, and $\sum \alpha_k = 1$. Assume $P = \sum \alpha_k P_k$ and $Q = \sum \alpha_k Q_k$ are φ-irreducible. Then if $\sigma(\mathcal{Q}_k - \mathcal{P}_k) \subset [0,\infty)$ for every k, then P efficiency dominates Q.*

Proof. By lemma 42, for every k we have for every $f \in L_0^2(\pi)$

$$\langle f, (\mathcal{Q}_k - \mathcal{P}_k)\, f\rangle \geq 0.$$

So, for every $f \in L_0^2(\pi)$,

$$\langle f, (\mathcal{Q} - \mathcal{P})\, f\rangle = \sum \alpha_k \langle f, (\mathcal{Q}_k - \mathcal{P}_k)\, f\rangle \geq 0.$$

So by lemma 42, $\sigma(\mathcal{Q} - \mathcal{P}) \subset [0,\infty)$, so by theorem 43, P effieciency dominates Q. □

8 Peskun Dominance

Definition (Peskun-Dominance)**.** *For two Markov kernels P and Q with stationary distribution π, P Peskun-dominates Q (or P dominates Q off the diagonal) if for π-almost every $x \in \mathbf{X}$,*

$$P(x, E - \{x\}) \geq Q(x, E - \{x\}), \quad \forall E \in \mathcal{F}.$$

Lemma 48. *(As seen in [10]) If P and Q are Markov kernels reversible wrt π, such that P Peskun-dominates Q, then $\mathcal{Q} - \mathcal{P}$ is a positive operator.*

Proof. Let $f \in L_0^2(\pi)$. Then

$$\begin{aligned}
\langle f, (\mathcal{Q} - \mathcal{P}) f \rangle &= \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(x) f(y) (Q(x, dy) - P(x, dy)) \pi(dx) \\
&= \frac{1}{2} \int_{x \in \mathbf{X}} f(x)^2 \pi(dx) - \frac{1}{2} \int_{y \in \mathbf{X}} f(y)^2 \pi(dy) \\
&\quad + \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(x) f(y) (Q(x, dy) - P(x, dy)) \pi(dx) \\
&= \frac{1}{2} \left[\int_{x \in \mathbf{X}} f(x)^2 \pi(dx) - \int_{y \in \mathbf{X}} f(y)^2 \pi(dy) \right. \\
&\quad -2 \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(x) f(y) (\delta_x(dy) + P(x, dy) - Q(x, dy)) \pi(dx) \\
&\quad \left. +2 \int_{y \in \mathbf{X}} f(y)^2 \pi(dy) \right] \\
&= \frac{1}{2} \left[\int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(x)^2 \left(\delta_x(dy) + P(x, dy) - Q(x, dy) \right) \pi(dx) \right. \\
&\quad -2 \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(x) f(y) \left(\delta_x(dy) + P(x, dy) - Q(x, dy) \right) \pi(dx) \\
&\quad \left. + \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} f(y)^2 (\delta_x(dy) + P(x, dy) - Q(x, dy)) \pi(dx) \right] \\
&= \frac{1}{2} \int_{x \in \mathbf{X}} \int_{y \in \mathbf{X}} (f(x) - f(y))^2 \left(\delta_x(dy) + P(x, dy) - Q(x, dy) \right) \pi(dx).
\end{aligned}$$

Notice that $(\delta_x(\cdot) + P(x, \cdot) - Q(x, \cdot))$ is a positive measure for π-almost every $x \in \mathbf{X}$, as for every π-almost every $x \in \mathbf{X}$, for every $E \in \mathcal{F}$, if $x \in E$, then as P

Peskun-dominates Q,

$$\begin{aligned}\delta_x(E) + P(x,E) - Q(x,E) &= 1 + P(x, E - \{x\}) - Q(x, E - \{x\}) \\ &\qquad\qquad\qquad + P(x,\{x\}) - Q(x,\{x\}) \\ &\geq 1 + Q(x, E - \{x\}) - Q(x, E - \{x\}) \\ &\qquad\qquad\qquad + P(x,\{x\}) - Q(x,\{x\}) \\ &\geq 1 + P(x,\{x\}) - Q(x,\{x\}) \\ &\geq P(x,\{x\}) \\ &\geq 0,\end{aligned}$$

and if $x \notin E$, then similarly

$$\delta_x(E) + P(x,E) - Q(x,E) = P(x, E - \{x\}) - Q(x, E - \{x\}) \geq 0.$$

□

Theorem 49. *If P and Q are φ-irreducible reversible Markov kernels reversible with respect to π, if P Peskun-dominates Q, then P efficiency dominates Q.*

Proof. For every $f \in L_0^2(\pi)$, by lemma 48, $\langle f, (\mathcal{Q}-\mathcal{P})f\rangle \geq 0$, so $\langle f, \mathcal{P}f\rangle \leq \langle f, \mathcal{Q}f\rangle$. Thus by theorem 41, P efficiency dominates Q. □

9 Non φ-Irreducible Kernels

In this section, we look at efficiency dominance for kernels that may not be φ-irreducible.

Lemma 50. *If P is a reversible Markov kernel reversible with respect to π, then* $\mathrm{null}(\mathcal{P} - \mathcal{I}) \subset L_0^2(\pi)$ *is closed.*

Proof. Let $f \in \overline{\mathrm{null}(\mathcal{P} - \mathcal{I})}$. Then there exists $\{f_n\} \subset \mathrm{null}(\mathcal{P} - \mathcal{I})$ such that $f_n \to f$ in the $L_0^2(\pi)$ norm, i.e. $\|f_n - f\| \to 0$. So, as $\mathcal{P}f_n = f_n$ for every $n \in \mathbf{N}$, we have

$$\|\mathcal{P}f - f\| = \|\mathcal{P}f - \mathcal{P}f_n + f_n - f\| \leq \|\mathcal{P}(f - f_n)\| + \|f_n - f\| \leq 2\,\|f_n - f\| \to 0,$$

so $\|\mathcal{P}f - f\| = 0$, and thus $\mathcal{P}f = f$ (as the equivalence class of functions equal π-almost everywhere), and $f \in \mathrm{null}(\mathcal{P} - \mathcal{I})$. □

Lemma 51. *If P and Q are reversible Markov kernels reversible with respect to π such that $\langle f, \mathcal{P}f\rangle \leq \langle f, \mathcal{Q}f\rangle$ for every $f \in L_0^2(\pi)$, then* $\mathrm{null}(\mathcal{P} - \mathcal{I}) \subset \mathrm{null}(\mathcal{Q} - \mathcal{I})$.

Proof. Let $g \in \mathrm{null}(\mathcal{P} - \mathcal{I})$. As $\mathcal{P}g = g$ and $\|\mathcal{Q}\| \leq 1$, by our hypothesis and the Cauchy-Schwartz inequality (Theorem 12.2 (1) in [7]),

$$\|g\|^2 = \langle g, \mathcal{P}g \rangle \leq \langle g, \mathcal{Q}g \rangle \leq \|\mathcal{Q}g\| \, \|g\| \leq \|\mathcal{Q}\| \, \|g\|^2 \leq \|g\|^2 ,$$

so $\|\mathcal{Q}g\| = \|g\|$.

Thus as $\mathcal{Q}$ is self-adjoint,

$$\begin{aligned}
\|\mathcal{Q}g - g\|^2 &= \|\mathcal{Q}g\|^2 - 2\langle g, Qg \rangle + \|g\|^2 \\
&\leq \|\mathcal{Q}g\|^2 - 2\langle g, \mathcal{P}g \rangle + \|g\|^2 \\
&= \|g\|^2 - 2\langle g, g \rangle + \|g\|^2 \\
&= \|g\|^2 - 2 \|g\|^2 + \|g\|^2 \\
&= 0,
\end{aligned}$$

so $\mathcal{Q}g = g$. □

Theorem 52. *If P and Q are reversible Markov kernels reversible with respect to π, such that for every $f \in L_0^2(\pi)$, $\langle f, \mathcal{P}f \rangle \leq \langle f, \mathcal{Q}f \rangle$, then P efficiency dominates Q.*

Proof. Let $f \in L_0^2(\pi)$. For every $\eta \in [0, 1)$, let $T_{\mathcal{P},\eta} = \mathcal{I} - \eta\mathcal{P}$ and $T_{\mathcal{Q},\eta} = \mathcal{I} - \eta\mathcal{Q}$. Then just as in theorem 41, $\sigma(T_{\mathcal{P},\eta}), \sigma(T_{\mathcal{Q},\eta}) \subset (0, \infty)$. Furthermore, notice that as $\langle f, \mathcal{P}f \rangle \leq \langle f, \mathcal{Q}f \rangle$ for every $f \in L_0^2(\pi)$,

$$\langle f, T_{\mathcal{P},\eta} f \rangle \geq \langle f, T_{\mathcal{Q},\eta} f \rangle, \quad \forall f \in L_0^2(\pi).$$

So by lemma 37,

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

Notice that as

$$\begin{aligned}
T_{\mathcal{P},\eta}^{-1} &= (\mathcal{I} - \eta\mathcal{P})^{-1} \\
&= \eta\mathcal{P} (\mathcal{I} - \eta\mathcal{P})^{-1} + (\mathcal{I} - \eta\mathcal{P}) (\mathcal{I} - \eta\mathcal{P})^{-1} \\
&= \eta\mathcal{P} (\mathcal{I} - \eta\mathcal{P})^{-1} + \mathcal{I},
\end{aligned}$$

and similarly $T_{\mathcal{Q},\eta}^{-1} = \eta\mathcal{Q} (\mathcal{I} - \eta\mathcal{Q})^{-1} + \mathcal{I}$, we have

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

If $E_{f,\mathcal{P}}(\{1\}) = 0$, then as $\eta \in [0,1)$ was arbitrary, by theorem 29, just as in theorem 41,

$$\begin{aligned}
v(f,P) &= \int_{\lambda\in\sigma(\mathcal{P})} \frac{1+\lambda}{1-\lambda} E_{f,\mathcal{P}}(d\lambda) \\
&= \int_{\lambda\in[-1,1)} \frac{1+\lambda}{1-\lambda} E_{f,\mathcal{P}}(d\lambda) \\
&= \lim_{\eta\to 1^-} \int_{\lambda\in[-1,1)} \frac{1+\lambda}{1-\eta\lambda} E_{f,\mathcal{P}}(d\lambda) \\
&= \|f\|^2 + 2 \lim_{\eta\to 1^-} \langle f, \mathcal{P}\left(\mathcal{I}-\eta\mathcal{P}\right)^{-1} f\rangle \\
&\le \|f\|^2 + 2 \lim_{\eta\to 1^-} \langle f, \mathcal{Q}\left(\mathcal{I}-\eta\mathcal{Q}\right)^{-1} f\rangle \\
&= \lim_{\eta\to 1^-} \int_{\lambda\in[-1,1)} \frac{1+\lambda}{1-\eta\lambda} E_{f,\mathcal{Q}}(d\lambda) \\
&= \int_{\lambda\in\sigma(\mathcal{Q})} \frac{1+\lambda}{1-\lambda} E_{f,\mathcal{Q}}(d\lambda) \\
&= v(f,Q).
\end{aligned}$$

(assuming $E_{f,\mathcal{Q}}(\{1\}) \neq 0$, as otherwise trivial).

If $E_{f,\mathcal{P}}(\{1\}) \neq 0$, then by Theorem 12.29 (b) in [7], 1 is an eigenvalue of $\mathcal{P}$. As $\mathrm{null}(\mathcal{P}-\mathcal{I})$ is a closed subspace by lemma 50 (proved above for this specific case, but more is true. The null space of every bounded operator is closed), we have by Theorem 12.4 in [7],

$$L_0^2(\pi) = \mathrm{null}(\mathcal{P}-\mathcal{I}) \bigoplus \mathrm{null}(\mathcal{P}-\mathcal{I})^{\perp}.$$

So, there exists $g \in \mathrm{null}(\mathcal{P}-\mathcal{I})$ and $f_0 \in \mathrm{null}(\mathcal{P}-\mathcal{I})^{\perp}$ such that $f = f_0 + g$.

As $g \in \text{null}(\mathcal{P} - \mathcal{I})$, $g \in \text{null}(\mathcal{Q} - \mathcal{I})$ by lemma 51. So by lemma 28, we have

$$\begin{aligned}
v(f, Q) &= \|f\|^2 + 2 \lim_{N\to\infty} \left[\sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \langle f, \mathcal{Q}^k f \rangle \right] \\
&= \|f\|^2 + 2 \lim_{N\to\infty} \left[\sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \langle f_0 + g, \mathcal{Q}^k (f_0 + g) \rangle \right] \\
&= \|f\|^2 + 2 \lim_{N\to\infty} \left[\sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \left(\langle f_0, \mathcal{Q}^k f_0 \rangle + 2\langle f_0, \mathcal{Q}^k g \rangle + \langle g, \mathcal{Q}^k g \rangle \right) \right] \\
&= \|f\|^2 + 2 \lim_{N\to\infty} \left[\sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \left(\langle f_0, \mathcal{Q}^k f_0 \rangle + 2\langle f_0, g \rangle + \langle g, g \rangle \right) \right] \\
&= \|f\|^2 + 2 \lim_{N\to\infty} \left[\sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \left(\langle f_0, \mathcal{Q}^k f_0 \rangle + \|g\|^2 \right) \right].
\end{aligned}$$

Note now that for every $N \in \mathbf{N}$, as $\mathcal{Q}$ is self-adjoint, by the Cauchy-Schwartz inequality,

$$
\begin{aligned}
&\sum_{k=1}^{N} \left(\frac{N-k}{N}\right) \langle f_0, \mathcal{Q}^k f_0 \rangle \\
&= \left(\frac{N-1}{N}\right) \langle f_0, \mathcal{Q} f_0 \rangle \\
&\quad + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{N-2m}{N}\right) \langle f_0, \mathcal{Q}^{2m} f_0 \rangle + \left(\frac{N-(2m+1)}{N}\right) \langle f_0, \mathcal{Q}^{2m+1} f_0 \rangle \right] \\
&= \left(\frac{N-1}{N}\right) \langle f_0, \mathcal{Q} f_0 \rangle \\
&\quad + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{N-2m}{N}\right) \|\mathcal{Q}^m f_0\|^2 + \left(\frac{N-(2m+1)}{N}\right) \langle \mathcal{Q}^m f_0, \mathcal{Q}^{m+1} f_0 \rangle \right] \\
&\geq \left(\frac{N-1}{N}\right) \langle f_0, \mathcal{Q} f_0 \rangle \\
&\quad + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{N-2m}{N}\right) \|\mathcal{Q}^m f_0\|^2 - \left(\frac{N-(2m+1)}{N}\right) \|\mathcal{Q}^m f_0\| \left\|\mathcal{Q}^{m+1} f_0\right\| \right] \\
&\geq \left(\frac{N-1}{N}\right) \langle f_0, \mathcal{Q} f_0 \rangle \\
&\quad + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{N-2m}{N}\right) \|\mathcal{Q}^m f_0\|^2 - \left(\frac{N-(2m+1)}{N}\right) \|\mathcal{Q}\| \|\mathcal{Q}^m f_0\| \|\mathcal{Q}^m f_0\| \right] \\
&\geq \left(\frac{N-1}{N}\right) \langle f_0, \mathcal{Q} f_0 \rangle \\
&\quad + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{N-2m}{N}\right) \|\mathcal{Q}^m f_0\|^2 - \left(\frac{N-(2m+1)}{N}\right) \|\mathcal{Q}^m f_0\|^2 \right] \\
&\geq \left(\frac{N-1}{N}\right) \langle f_0, \mathcal{Q} f_0 \rangle + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{1}{N}\right) \|\mathcal{Q}^m f_0\|^2 \right].
\end{aligned}
$$

So, plugging this into what we had earlier,

$$
\begin{aligned}
v(f, Q) &= \|f\|^2 + 2 \lim_{N\to\infty} \left[\sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \left(\langle f_0, \mathcal{Q}^k f_0 \rangle + \|g\|^2 \right) \right] \\
&\geq 2 \lim_{N\to\infty} \left[\left(\frac{N-1}{N} \right) \langle f_0, \mathcal{Q} f_0 \rangle + \sum_{m=1}^{\lfloor N \rfloor} \left[\left(\frac{1}{N} \right) \|\mathcal{Q}^m f_0\|^2 \right] \right. \\
&\qquad \left. + \sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \|g\|^2 \right] \\
&\geq 2 \lim_{N\to\infty} \left[\left(\frac{N-1}{N} \right) \langle f_0, \mathcal{Q} f_0 \rangle + \sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \|g\|^2 \right] \\
&= 2 \lim_{N\to\infty} \left(\frac{N-1}{N} \right) \langle f_0, \mathcal{Q} f_0 \rangle + 2 \lim_{N\to\infty} \sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \|g\|^2 \\
&= 2 \langle f_0, \mathcal{Q} f_0 \rangle + 2 \lim_{N\to\infty} \sum_{k=1}^{N} \left(\frac{N-k}{N} \right) \|g\|^2 .
\end{aligned}
$$

Finally, notice that as $\lim_{N\to\infty} \sum_{k=1}^{N} \left(\frac{N-k}{N}\right) \|g\|^2 = \infty$, we have $v(f, Q) = \infty$, so $v(f, P) \leq \infty = v(f, Q)$. $\square$

With theorem 52, we can establish many of the same results on efficiency dominance for more general kernels just as we did in previous sections for φ-irreducible kernels.

List of Results that carry over for (not necessarily φ-irreducibe) reversible kernels:

- proposition 31
- if direction of theorem 41, possibly converse as well
- if direction of theorem 43, possibly converse as well
- theorem 44 if converse of theorem 52 is true, if not then it is false (I think, doubt the antisymmetric)
- theorem 47
- theorem 49 (as seen in [10])

Notes

- Not many Markov Chains have compact Markov operators. This is said in [5], bottom of page 7. They cite a source for this. It may be worth showing some examples.
- [5] cite a source that identifies the set of minimal variance operators when in matrix form.

References

[1] Efficiency of Reversible MCMC Methods: Elementary Derivations and Applications to Composite Methods, Neal and Rosenthal

[2] *Real Analysis, 2nd Ed.*, Folland

[3] General State Space Markov Chains and MCMC Algorithms, Roberts and Rosenthal

[4] Rates of Convergence for Everywhere-Positive Markov Chains, Baxter and Rosenthal

[5] Ordering Monte Carlo Markov Chains, Mira and Geyer

[6] *A Course in Functional Analysis*, Conway

[7] *Functional Analysis, 2nd Ed.*, Rudin

[8] Variance Bounding Markov Chains, Rosenthal and Roberts

[9] *Topology, 2nd Ed.*, Munkres

[10] A Note on Metropolis-Hastings Kernels for General State Spaces, Tierney

[11] Equivalences of Geometric Ergodicity of Markov Chains, Gallegos-Herrada and Ledvinka and Rosenthal

[12] *Markov Chains and Stochastic Stability*, Meyn and Tweedie

[13] *Markov Chains*, Douc and Moulines and Priouret and Soulier

[14] Convergence Rates and Efficiency Dominance of Markov Chains, Salmon