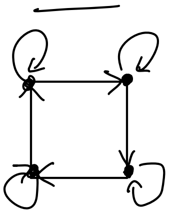
Spectral Gap of a Nonreversible Chain

Ex: Consider the random walk on $\mathbb{Z}/N\mathbb{Z}$, with transitions given by jumps of size a_i with prob p_i, $i=1,\dots,k$.

1) If $k=2$, $a_1=1$, $a_1=-1 \pmod N$, and $p_1=p_2=1/2$, this is just a simple random walk on the N-cycle. It can be shown (using results from today) that the relaxation time of this walk is $O(N^2)$.

2) What if $k=2$, $a_1=0$, $a_2=1$, and $p_1=p_2=1/2$?

N=4



This is a lazy chain that always moves right, if it moves at all. We will see that empirical averages of statistics of this chain converge in time $O(N)$, even though it mixes in time $O(N^2)$.

In the first case, the walk is reversible, and in the second, it is not. Nonetheless, there is an elegant underlying theory, introduced by Sourav Chatterjee, that unifies both settings!

Preliminaries

We restrict ourselves to time-homogeneous Markov chains $\{X_n\}_{n\in\mathbb{N}}$ on S, a discrete state space, with $|S|=d$.

Throughout, P is the transition matrix and

$$L := I - P$$

is the generator. If μ is an invariant measure such that $\mu(x)>0$ for all $x\in S$, then we denote the induced inner product by

$$\langle f, g \rangle := \sum_{x \in S} f(x) g(x) \mu(x),$$

with induced norm $\|f\|$.

<u>def</u>: The <u>spectral gap</u> of such a chain is

$$\gamma = \gamma(L) := \sigma_2(L),$$

where $0 \leq \sigma_1(L) \leq \sigma_2(L) \leq \dots \leq \sigma_d(L)$ are the singular values of L wrt. the inner product induced by μ on $\mathbb{C}^d$, with multiplicity.

The <u>relaxation time</u> is then $\tau = \tau(L) := 1/\gamma$.

<u>Rmk</u>: If the chain is reversible, the eigenvalues of P are arranged as $1 = \lambda_1 \geq \lambda_2 \dots \geq \lambda_d \geq -1$, with spectral gap

$$\gamma_{rev} = 1 - \lambda_2(P) = \sigma_2(L).$$

If the chain is lazy, so that its relaxation time is the inverse of $1 - \max\{\lambda_2, |\lambda_d|\} = 1 - \lambda_2$, then Chatterjee's relaxation time is the same as the reversible version too (c.f. LPW, Ch. 13).

As in the reversible case, the spectral gap yields very useful information about mixing. First, however, we survey some previous attempts at constructing such a framework for non-reversible chains.

<u>Prior Work</u>

One may consider <u>reversibilizing</u> chains:

1) Additive: take a chain with transitions

$$A = \tfrac{1}{2}(P + P^*)$$ adjoint wrt. μ inner product!

2) Multiplicative: take a chain with transitions:

$$M = P P^*.$$

This idea, originally due to J. A. Fill (1991), giving upper bounds on mixing times for nonreversible chains.

One can also analyze the spectrum of P as a subset of $\mathbb{C}$ or the singular values of P itself, but neither has a clear connection to the singular values of the generator or very strong bounds on mixing times for all nonreversible chains. Of course, fruitful analysis of nonreversible chains can be done without using any notion of a spectral gap.

↳ See, for instance, Diaconis, Holmes, and Neal (2000).

Main Result: Empirical Averages

Let μg be denote the μ-average of $g : S \to \mathbb{R}$,

$$\mu g := \sum_{x \in S} g(x) \mu(x) = \mathbb{E}_{\mu}[g].$$

Given $n \in \mathbb{N}$, denote the nth empirical average by

$$\mu_n g := \frac{1}{n} \sum_{i=0}^{n-1} g(X_i).$$

steps of nonreversible chain

If $X_0 \sim \mu$ is stationary, then we consider

$$\Delta_n := \sup_{g : \|g - \mu g\| = 1} \|\mu_n g - \mu g\|_{L^2},$$

the worst case concentration of $\mu_n g$ about its mean μg.

Thm: (Theorem 1.2 in Chatterjee)

For any $n \in \mathbb{N}$, $\Delta_n \leq \sqrt{\frac{4\tau}{n}}$, while for any $n \leq \tau/3$,

$$\Delta_n \geq \frac{1}{132},$$

and for all $n \in \mathbb{N}$,

$$\max_{n \le k \le 3n} \Delta_k \ge \frac{\tau}{2n + 3\tau}.$$

For reversible chains,

$$\sup_{g : \|g - \mu g\| = 1} \|P^n g - \mu g\|_{L^2} = \left(1 - \frac{1}{T_{rel}}\right)^n,$$

and this same quantity characterizes decorrelation times for rev. chains.

For non-reversible chains, this is not the case.

Ex: Returning to the right-moving 1/2-lazy walk on the N-cycle, the relaxation time is $O(N)$, but

$$X_n \equiv X_0 + \frac{n}{2} + O(\sqrt{n}) \pmod{N}$$

doesn't decorrelate from X_0 until time N^2, roughly speaking.

Main Result: Mixing Times

Take

$$P^n_x = \delta_x P^n$$

$$T_{mix}(\varepsilon) := \inf\{n \ge 0 : \|P^n_x - \mu\|_{TV} \le \varepsilon\}.$$

For reversible chains

$$(T_{rel} - 1)\log\frac{1}{2\varepsilon} \le T_{mix}(\varepsilon) \le T_{rel} \cdot \log\frac{1}{\varepsilon \mu_{min}},$$

where $\mu_{min} = \min_{x \in S} \mu(x)$ (> 0, as our chains are positive recurrent), as shown in LPW, Ch. 12. For nonreversible chains, a similar result holds.

Thm: (Theorem 1.4 in Chatterjee)

For any $\varepsilon \in (0, 1/5)$,

$$\tau \le \left(\frac{4}{\log\left(\frac{2}{1 + 4\sqrt{\varepsilon} + \varepsilon^2}\right)} + 2\right) T_{mix}(\varepsilon),$$

and if $P(x,x) \geq 1/2$ for all $x \in S$, then

$$\tau_{mix}(\varepsilon) \leq 1 + 4\tau^2 \log\left(\frac{1}{2\varepsilon \mu_{min}}\right).$$

Rmk: The τ^2 upper bound is necessary, as seen in our running example of the right-moving lazy chain on the cycle.

Finally, for normal P, this analysis nicely generalizes the reversible setting. Indeed, P then admits a complex diagonalization

$$P = \mathcal{U} D \mathcal{U}^*, \quad D = \mathrm{diag}(\lambda_0, \dots, \lambda_{d-1}),$$

so we may write

$$L = \mathcal{U}(I-D)\mathcal{U}^*, \quad L^*L = \mathcal{U}(I-D)^*(I-D)\mathcal{U}^*.$$

Thus, the singular values of L are $0 = |1-\lambda_0| \leq |1-\lambda_1| \leq \dots \leq |1-\lambda_{d-1}|$, and $\gamma = |1-\lambda_1|$ in the normal (and hence reversible) case.

Ex: Returning to the walk on $\mathbb{Z}/N\mathbb{Z}$ with jumps a_i w.p. p_i, the transition matrix is circulant:

$$P = \begin{pmatrix} 0 & p_1 & p_2 & p_3 \\ p_3 & 0 & p_1 & p_2 \\ p_2 & p_3 & 0 & p_1 \\ p_1 & p_2 & p_3 & 0 \end{pmatrix},$$

which one can verify satisfies $PP^* = P^*P$, i.e., P is normal! The spectrum of a circulant matrix is also easy to characterize explicitly, with eigenvalues

$$\lambda_j = \sum_{r=1}^{k} p_r e^{2\pi i j a_r / N},$$

which can be seen by diagonalizing the circulant matrix via a discrete Fourier transform [e.g., eigenvectors are $(1, \omega_j, \omega_j^2, \ldots, \omega_j^{N-1})$ for the Nth roots of unity $\omega_j = \exp(2\pi i j/N)$]. Hence,

$$\tau = \max_{1 \le j \le N-1} \left| 1 - \sum_{r=1}^{k} p_r e^{2\pi i j a_r/N} \right|^{-1}.$$

1) If $a_1 = 1$, $a_2 = -1$, $p_1 = p_2 = 1/2$, then

$$\tau = \max_j \frac{1}{1-\cos(2\pi j/N)} = \frac{1}{1-\cos(2\pi/N)} = O(N^2).$$

2) If $a_1 = 0$, $a_2 = 1$, $p_1 = p_2 = 1/2$, then

$$\tau = \max_j \left| 1 - \tfrac{1}{2}(1 + e^{2\pi i j/N}) \right|^{-1}$$

$$= \max_j \frac{1}{|\sin(\pi j/N)|}$$

$$\frac{1}{2} - \frac{1}{2} e^{2\pi i j/N} = \frac{1 - e^{2\pi i j/N}}{2} = \frac{e^{-\pi i j/N} - e^{\pi i j/N}}{2e^{-\pi i j/N}} = \frac{-i\sin(\pi j/N)}{e^{-\pi i j/N}}$$

$$= \frac{1}{\sin(\pi/N)}$$

$$= O(N),$$

confirming the example from the start!

Of course, we can also come up with more interesting examples.

Ex: (Chung – Diaconis – Graham, 1986)
Again, on $\mathbb{Z}/N\mathbb{Z}$, take

$$X_n = 2X_{n-1} + \varepsilon_n \pmod N,$$

Rmk: Second largest singular value of P is $1 - \Theta(\frac{1}{N^2})$, so sing. vals of P do not capture mixing!

where $\varepsilon_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{-1, 0, 1\})$. This is a nonreversible chain with mixing time $O(\log N \log\log N)$. It is also a classic example of the cutoff phenomenon, which occurs at a multiple of $\log N$ for almost all odd N [c.f. Eberhard + Varjú (2021)]. Chatterjee shows that for all prime $N \geq 3$, $\tau = \Theta(\log N)$, confirming that empirical averages also take this long to converge.

Other examples include card-shuffling à la Diaconis, random walks on groups, and a local chain on the d-dimensional torus with some number-theoretic intricacies.

Main Technical Arguments

To prove the stated results, most arguments are purely linear algebraic. Here, we provide a sample of the arguments that Chatterjee uses to connect his spectral gap to mixing times. First, we have the following result, whose proof is entirely linear algebra.

Thm: (Theorem 1.3 in Chatterjee)

Let γ be the spectral gap of a (nonreversible) chain w/ transition matrix P, and let γ_M, γ_A be the spectral gaps of its multiplicative and additive reversibilizations, resp. Then, $\frac{1}{2}\gamma_A \leq \gamma \leq \sqrt{2\gamma_A}$, $\frac{1}{2}\gamma_M \leq \gamma$, and if $P(x,x) \geq 1/2$ for all $x \in S$, $\gamma \leq \sqrt{2\gamma_M}$.

pf: (Sketch)

If λ_M is the second largest eigenvalue of M, then $\gamma_M = 1 - \lambda_M$, and the second largest singular value of P is given by $\sqrt{\lambda_M}$. Hence, if $f: S \to \mathbb{C}$ is s.t. $\mu f = 0$,

$$\|(I-P)f\| \geq \|f\| - \|Pf\|$$
$$\geq (1-\sqrt{\lambda_M})\|f\|,$$

so that $\gamma \geq 1-\sqrt{\lambda_M} = 1-\sqrt{1-\gamma_M}$ $\quad \sqrt{1-x} \leq 1-\frac{1}{2}x$
$$\geq 1-(1-\tfrac{1}{2}\gamma_M) \quad \text{for } x \leq 1.$$
$$= \frac{\gamma_M}{2},$$

with similar reasoning in the case of γ_A. For the upper bounds on γ, we instead observe that if $L = I - P$,

$L_A = I - A = I - \frac{1}{2}(P + P^*)$, $L_M = I - M = I - PP^*$,

then

$$LL^* = I - P - P^* + PP^*$$
$$= 2I - P - P^* - (I - PP^*)$$
$$= 2L_A - L_M.$$

Hence, if $\mu f = 0$ and $\|f\| = 1$, then

$$\langle f, LL^* f\rangle = \overset{(1)}{\langle 2f, L_A f\rangle} - \overset{(2)}{\langle f, L_M f\rangle}.$$

As L_M is PSD wrt $\langle \cdot, \cdot \rangle_\mu$, this is bounded above by (1), with $(2) \geq 0$. Taking the inf. over all such f yields

$$\gamma^2 \leq 2\gamma_A.$$

Similar considerations show that $\lambda_2(M) \leq \lambda_2(A)$, so that $\gamma_M \geq \gamma_A \implies \gamma_M \geq \frac{1}{2}\gamma^2$ as well. $\square$

Equipped with this fact, we can prove the mixing time result stated above. First, we have the following technical lemma, proven below for completeness.

<u>Lemma 1:</u> If P, Q are Markov transition kernels on S with common invariant measure μ, and $R := PQ$, then

$$\max_{x \in S} \|R_x - \mu\|_{TV} \le 2 \max_{x \in S} \|P_x - \mu\|_{TV} \cdot \max_{x \in S} \|Q_x - \mu\|_{TV}.$$

pf: Consider $M \in \mathbb{R}^{d \times d}$, with entries $M(x,y) = \mu(y)$. One can verify that $M^2 = M$, and $MP = PM = M$, $MQ = QM = M$. In turn,

$$R - M = PQ - M = (P - M)(Q - M).$$

Hence,

$$\max_{x \in S} \|\overset{\delta_x R}{R_x} - \mu\|_{TV} = \frac{1}{2} \max_{x \in S} \sum_{y \in S} |R(x,y) - \mu(y)|$$

$$= \frac{1}{2} \|R - M\|_{\ell^\infty \to \ell^\infty}$$

$$\le \frac{1}{2} \|P - M\|_{\ell^\infty \to \ell^\infty} \|Q - M\|_{\ell^\infty \to \ell^\infty}$$

$$= 2 \max_{x \in S} \|P_x - \mu\|_{TV} \cdot \max_{x \in S} \|Q_x - \mu\|_{TV}.$$

$\square$

We can now prove the main result concerning mixing times of nonreversible chains.

pf: Take $f : S \to \mathbb{C}$ s.t. $\mu f = 0$, and fix some

$\varepsilon \in (0, 1/5)$, denoting $n := \tau_{mix}(\varepsilon)$. Assume that τ' is the relaxation time of P^n, i.e.,

$$\|f - P^n f\| \geq \frac{\|f\|}{\tau'}.$$

Now, P is an L^2-contraction (row sums are one, so spectral radius is at most one), so

$$\|f - P^n f\| \leq \sum_{k=0}^{n-1} \overbrace{\|P^k f - P^{k+1} f\|}^{\|P^k(I-P)f\| \leq \|f - Pf\|}$$

$$\leq n \|f - Pf\|,$$

yielding $\|f - Pf\| \geq \frac{\|f\|}{n\tau'}$, i.e., $\gamma \geq \frac{1}{n\tau'} \Rightarrow \tau = \frac{1}{\gamma} \leq n\tau'$.

If A^n is the additive reversibilization of P^n,

$$A^n := \frac{1}{2}\left((P^n)^* + P^n\right),$$

then

$$(A^n)^2 = \frac{1}{4}\left[(P^*)^{2n} + (P^n)^* P^n + P^n (P^n)^* + P^{2n}\right]$$

so that

$$\|(A^n)^2_x - \mu\|_{TV} \leq \frac{1}{4}\Big[\overbrace{\|(P^*)^{2n}_x - \mu\|_{TV}}^{\leq 1} + \overbrace{\|(P^n)^* P^n_x - \mu\|_{TV}}^{\leq 2\varepsilon}$$
$$+ \underbrace{\|P^n (P^n)^*_x - \mu\|_{TV}}_{\leq 2\varepsilon} + \underbrace{\|P^{2n}_x - \mu\|_{TV}}_{\leq 2\varepsilon^2}\Big],$$

and by the technical lemma, taking the max. over $x \in S$ gives

$$\max_{x \in S} \|(A^n)^2_x - \mu\|_{TV} \leq \frac{1}{4}(1 + 4\varepsilon + 2\varepsilon^2) =: \varepsilon' < 1/2,$$

as $\varepsilon \in (0, 1/5)$. If σ_{rel}, $\sigma_{mix}(\cdot)$ are the relaxation / mixing times of the chain defined by A, this shows that $\tau_{mix}(\varepsilon') \le 2$. Because this chain is reversible,

$$\tau_{rel} \le \frac{2}{\log\left(\frac{1}{2\varepsilon'}\right)} + 1,$$

while $\tau' \le 2\tau_{rel}$ by the relationship between the additive reversibilization and original chain from above. Hence,

$$\tau \le 2\tau_{mix}(\varepsilon)\left(\frac{2}{\log\left(\frac{1}{2}(1+2\varepsilon^2+4\varepsilon)\right)} + 1\right).$$

On the other hand, if $P(x,x) \ge 1/2$, then for $M = PP^*$, we have that $\gamma \le \sqrt{2\gamma_M}$. Hence, if $\mu f = 0$,

$$\|Pf\|^2 = \langle f, \overset{\sigma(P^*P) = \sigma(PP^*)}{P^*Pf}\rangle \le (1-\gamma_M)\|f\|^2$$
$$\le \left(1 - \frac{\gamma^2}{2}\right)\|f\|^2.$$

For all $n \in \mathbb{N}$, in fact, $\|P^n f\| \le \left(1-\frac{\gamma^2}{2}\right)^{n/2}\|f\|$. Now, we compute

$$|\mathbb{E}_x(f(X_n)) - \mu f| \le \frac{1}{\mu(x)} \sum_{y \in S} \mu(y)\,|\mathbb{E}_y(f(X_n)) - \mu f|$$

$$(\text{Cauchy–Schwarz}) \qquad \le \frac{1}{\mu(x)}\sqrt{\sum_{y\in S}\mu(y)\,|\mathbb{E}_y(f(X_n)) - \mu f|^2}$$

$$= \frac{1}{\mu(x)}\|P_n f - \mu f\|.$$

Taking the supremum over $f: S \to [-1,1]$, we obtain

$$\|P_x^n - \mu f\|_{TV} \le \frac{1}{2\mu(x)}\left(1 - \frac{\gamma^2}{2}\right)^{n/2}.$$

Now, if $\tau_{mix}(\varepsilon) = 0$, we are done. If $\tau_{mix}(\varepsilon) \ge 1$, then for

$n = \tau_{mix}(\varepsilon) - 1$, we have, for some $x \in S$

$$\left(1 - \frac{\gamma^2}{2}\right)^{(\tau_{mix}(\varepsilon)-1)/2} > 2\varepsilon\mu(x) \geq 2\varepsilon\mu_{min},$$

which is strictly pos. by assumption. To conclude,

$$\frac{\tau_{mix}(\varepsilon) - 1}{2} \leq \frac{\log(2\varepsilon\mu_{min})}{\log\left(1 - \frac{\gamma^2}{2}\right)}$$

$$[\log(1-x) \leq -x] \qquad \leq \frac{2}{\gamma^2}\log\left(\frac{1}{2\varepsilon\mu_{min}}\right). \qquad \square$$

pf: (Empirical Averages)

If $\tau = \infty$, then $\gamma = 0$, so we can choose $g : S \to \mathbb{R}$ with $\mu g = 0$, $\|g\| = 1$, an $Lg = 0 \Rightarrow g = Pg$, in turn, from

$$\Delta_n^2 \geq \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=0}^{n-1} g(X_i)\right)^2\right]$$

and the fact that $g = P^n g$ for all $n \in \mathbb{N}$, the lower bounds follow.

If $\tau < \infty$, then 0 is a singular value of L with mult. 1. In particular, $\dim \operatorname{Im}(L) \geq d - 1$. But

$$\dim \operatorname{Im}(L) \subseteq \underbrace{\operatorname{span}\{\mu\}^\perp}_{\text{dim. } d-1}$$

so $\operatorname{Im}(L) = \operatorname{span}\{\mu\}^\perp$. Given any $g : S \to \mathbb{R}$, we can hence find f s.t. $Lf = g - \mu g$, and

$$\gamma\|f\| \leq \|Lf\| = \|g - \mu g\|.$$

Claim: $\sum_x \mu(x)\left[\mathbb{E}_x(\mu_n g - \mu g))^2\right] \leq \frac{4\tau^2\|g-\mu g\|^2}{n^2}$.

pf: Note that, for f, g as above

$$\mathbb{E}_x\left(g(X_n) - \mu g\right) = \mathbb{E}_x\left(f(X_n) - P f(X_n)\right)$$

$$= \mathbb{E}_x\left(f(X_n) - f(X_{n+1})\right),$$

so

$$\mathbb{E}_x(\mu_n g - \mu g) = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_x\left(g(X_k) - \mu g\right)$$

$$= \frac{1}{n}\left(f(x) - \mathbb{E}_x f(X_n)\right).$$

Hence,

$$\sum_x \mu(x)\left[\mathbb{E}_x(\mu_n g - \mu g))^2\right] = \frac{1}{n^2}\sum_x \mu(x)\left(\mathbb{E}_x f(X_n) - f(x)\right)^2$$

[Jensen] $\leq \frac{1}{n^2}\sum_x \mu(x)\, \mathbb{E}_x\left[\left(f(X_n) - f(x)\right)^2\right]$

[AM-GM] $\leq \frac{1}{n^2}\sum_x \mu(x)\left[2\mathbb{E}_x f(X_n)^2 + 2f(x)^2\right]$

$\left[\mathbb{E}_x f(X_n)^2 = P^n f(x)^2\right]$ $= \frac{4\|f\|^2}{n^2}$

$\left[\|f\| \leq \tau\|g - \mu g\|\right]$ $\leq \frac{4\tau^2\|g-\mu g\|^2}{n^2}$.

Now, given $g: S \to \mathbb{R}$ with $\mu g = 0$, $\|g\| = 1$. We find that

$$\Delta_n^2 = \sum_x \mu(x) \mathbb{E}_x \left[\left(\frac{1}{n} \sum_{i=0}^{n-1} g(X_i) \right)^2 \right]$$

$$= -\frac{1}{n^2} \sum_x \sum_{i=0}^{n-1} \mu(x) \mathbb{E}_x [g(X_i)^2] + \frac{2}{n^2} \sum_x \sum_{i=0}^{n-1} \mu(x) \mathbb{E}_x \left[g(X_i) \sum_{j=i}^{n-1} g(X_j) \right]$$

$$= -\frac{\|g\|^2}{n} + \frac{2}{n^2} \sum_x \sum_{i=0}^{n-1} \mu(x) \mathbb{E}_x \left[g(X_i) \sum_{j=i}^{n-1} g(X_j) \right].$$

Now, for any $i \in [n-1]$,

$$h(x) := \mathbb{E} \left[\sum_{j=i}^{n-1} g(X_j) \mid X_i = x \right]$$

$$= \mathbb{E} \left[\sum_{j=0}^{n-i-1} g(X_j) \mid X_0 = x \right]$$

$$= (n-i) \mathbb{E}_x [\mu_{n-i} g].$$

Then,

$$\mathbb{E}_x \left[g(X_i) \sum_{j=i}^{n-1} g(X_j) \right] \le \left(\mathbb{E}_x [g(X_i)^2] \mathbb{E}[h(X_i)^2] \right)^{1/2},$$

so by Cauchy-Schwarz,

$$\sum_x \mu(x) \mathbb{E}_x \left[g(X_i) \sum_{j=i}^{n-1} g(X_j) \right] \le \|g\| \|h\|$$

$$\leq \|g\| \sqrt{(n-i)^2 \sum_x \mu(x) \mathbb{E}_x(\mu_{n-i} g)^2}$$

$$\leq \|g\| \cdot \sqrt{4 \tau^2 \|g\|^2}$$

$$= 2\tau \|g\|^2.$$

Together, this yields

$$\Delta_n^2 \leq \frac{4\tau \|g\|^2}{n} - \frac{\|g\|^2}{n} \leq \frac{4\tau^2}{n}$$

for all g with $\mu g = 0$, $\|g\| = 1$, so $\Delta_n \leq \sqrt{\frac{4\tau}{n}}$.

To prove the lower bounds, we need the following:

<u>Claim:</u> If $n = n_1 + \dots + n_k$, then

$$n \Delta_n \leq n_1 \Delta_{n_1} + \dots + n_k \Delta_{n_k},$$

and $\Delta_n \leq 1$ for all n.

<u>pf:</u> Take any $g : S \to \mathbb{R}$ w/ $\|g - \mu g\| = 1$. WLOG, take $\mu g = 0$. Then, by stationarity

$$n \|\mu_n g - \mu g\| = \left\| \sum_{i=0}^{n-1} g(X_i) \right\|_{L^2}$$

$$\leq \left\| \sum_{i=0}^{n_1-1} g(X_i) \right\|_{L^2} + \dots + \left\| \sum_{i=n_1+\dots+n_{k-1}}^{n-1} g(X_i) \right\|_{L^2}$$

$$\leq n_1 \Delta_{n_1} + \dots + n_k \Delta_{n_k},$$

and $n = 1 + \dots + 1$ proves the last claim, as $\Delta_1 = 1$.

Now, we aim to witness the lower bounds. Suppose f minimizes $\|Lf\|$ subject to $\mu f = 0$, $\|f\| = 1$. Because L has real entries, so does

f. Also, $r = \|Lf\|$. For $g = Lf$, let:

$$nr \cdot \frac{1}{n^2}\sum_{k=n}^{2n-1}\sum_{m=0}^{k-1} P^m\left(\frac{g}{\|g\|}\right),$$

$$u_n(x) := f(x) - \sum_{k=n}^{2n-1} \mathbb{E}_x f(X_k),$$

so if $n \asymp \frac{1}{r}$ and $g/\|g\|$ averages are small, so is $\|u_n\|$

$$v_n(x) := \frac{1}{n}\sum_{k=n}^{2n-1} \mathbb{E}_x f(X_k),$$

$\|v_n\|$ is small if time averages of f are small.

$$w_n(x) := \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_x\, g(X_k).$$

Note that $1 = \|f\| \le \|u_n\| + \|v_n\|$, and $\mu g = 0$, and

$$u_n(x) = \frac{1}{n}\sum_{k=n}^{2n-1}\left(f(x) - \mathbb{E}_x f(X_k)\right)$$

$$= \frac{1}{n}\sum_{k=n}^{2n-1} k\, w_k(x),$$

as $\mathbb{E}_x \mu_k g = \frac{1}{k}\left(f(x) - \mathbb{E}_x f(X_k)\right)$ for solutions to $Lf = g - \mu g$.

As $\|w_k\| \le \|g\|\Delta_k = r\Delta_k$, we obtain

$$\|u_n\| \le \frac{1}{n}\sum_{k=n}^{2n-1} k\|w_k\| \le \frac{r}{n}\sum_{k=n}^{2n-1} k\,\Delta_k$$

$$\le 2nr \max_{n \le k \le 2n-1} \Delta_k.$$

Conversely,

$$\|v_n\| = \frac{1}{n}\left\| 2n\,\mu_{2n} f - n\,\mu_n f\right\|$$

$$\le 2\|\mu_{2n} f\| + \|\mu_n f\|$$

$$\le 2\Delta_{2n} + \Delta_n, \quad [\|f - \mu f\| = 1]$$

so $1 \le (2nr + 3) \max_{n \le k \le 2n} \Delta_k$, i.e.,

$$\max_{n \le k \le 2n} \Delta_k \ge \frac{\tau}{2n+3\tau}.$$

Now, take $m \le \tau/3$, and take $n = 12m$. There is $k \in \{n, \dots, 2n\}$ such that

$$\Delta_k \ge \frac{1}{2n\gamma+3} = \frac{1}{24m\gamma+3}$$

$$\ge \frac{1}{8\tau\gamma+3}$$

$$= \frac{1}{11}.$$

If $k = qm + r$, then

$$\frac{1}{11} \le \Delta_k \le \frac{qm}{k}\Delta_m + \frac{r}{k}\Delta_r \le \Delta_m + \frac{r}{k}$$

$$\le \Delta_m + \frac{1}{12},$$

so $\Delta_m \ge \frac{1}{132}$. $\square$