

SINGULAR-VALUE GAP OF NONREVERSIBLE MARKOV PROCESSES

Ruochuan Xu

Stanford University

ruochuan@stanford.edu

May 2026

ABSTRACT. We consider a generalization of the spectral gap of reversible Markov generators to non-reversible processes, following the recent work [Cha25] on nonreversible finite-state Markov chains. Extending Chatterjee’s observations, we find that this spectral quantity that we call the *singular-value gap* characterizes the convergence of empirical averages, providing upper and lower bounds for finite-time variance uniformly over L^2 -functions. A key observation is that when the singular-value gap is positive, the generator is invertible on the L^2 -orthogonal complement of constant functions. In particular, the Poisson equation $-Lf = g$ can be solved, which enables our proof and connects our results to asymptotic variance and associated central limit theorems.

We also compare the singular-value gap with the spectral gap of the reversibilized process, the mixing time in total-variation distance, and the Cheeger constant. Several examples are provided throughout the text. Among other potential applications of the singular-value gap, these examples illustrate that a positive singular-value gap can help with variance reduction for observable classes in MCMC sampling, uncover slow-mixing mechanisms, and certify convergence of empirical averages for diffusion operators with complicated spectrum.

CONTENTS

1. Introduction	3
1.1. Setting and motivation	3
1.2. Main results	6
1.3. Related literature	8
2. Singular-value gap controls empirical variance	9
2.1. Local singular-value gap	10
2.2. Global singular-value gap	14
2.3. Sharpness of upper and lower bounds	17
2.4. Sharpened bounds with access to asymptotic variance	19
3. Ornstein–Uhlenbeck process with rotation	22
3.1. Explicit variance computation	22
3.2. Uniform acceleration for nonradial observables	23
3.3. Gaussian chaos and sharpness of variance bounds	24
3.4. Practical implication for variance reduction	25
4. Relation to other mixing quantities	27
4.1. Relation to the reversibilized process	27
4.2. Relation to the mixing time	29
4.3. Relation to the Cheeger constant	34
5. Applications	38
5.1. Shift-register chain on binary de Bruijn graph	38
5.2. A hypoelliptic shear on the torus	40
Appendix A. Omitted details	42
A.1. The Poisson-equation formula for asymptotic variance	42
A.2. Eigenfunctions of the rotating Ornstein–Uhlenbeck generator	43
A.3. Spectrum of the Jacobi operator for the reflected birth-death chain	44
A.4. Compactness for the hypoelliptic shear on the torus	46
Acknowledgements	48
References	49

1. INTRODUCTION

1.1. Setting and motivation. Let $(X_t)_{t \geq 0}$ be a time-homogeneous Markov process with an invariant probability measure μ on a measurable space S . The associated semigroup $(P_t)_{t \geq 0}$ on $L^2(\mu)$ is given by $P_t g(x) = \mathbb{E}_x[g(X_t)]$, where $\mathbb{E}_x$ denotes the expectation conditioned on $X_0 = x$. Assume that $(P_t)_{t \geq 0}$ is strongly continuous, with the generator L defined on $\text{Dom}(L)$ by $Lg = \lim_{t \downarrow 0} \frac{P_t g - g}{t}$, where $\text{Dom}(L)$ consists of L^2 -functions g for which the limit exists. By the Hille-Yosida theorem, L is closed and $\text{Dom}(L)$ is dense in $L^2(\mu)$. We sometimes consider discrete-time Markov chains as well, in which case the setting will be clarified. Let

$$\langle f, g \rangle := \int \bar{f} g \, d\mu$$

denote the $L^2(\mu)$ inner product, and $\|\cdot\|$ the corresponding norm. The empirical average $\mu_t g$ defined below can be understood as a Bochner integral, since $s \mapsto g(X_s)$ is continuous as an $L^2(\mathbb{P}_\mu)$ -valued function by the strong continuity of $(P_t)_{t \geq 0}$. If the process has jointly measurable sample paths, this coincides with the usual pathwise time integral; for simplicity, we assume joint-measurability. We shall also restrict attention to centered observables g in $L_0^2(\mu) := \{g \in L^2(\mu) : \int g \, d\mu = 0\}$. Note that for g in $L^2(\mu)$, we can always take $g - \mu g \in L_0^2(\mu)$.

A basic problem in the study of a Markov process is to quantify the speed with which the process converges. Two natural but distinct questions are: how quickly the time- t law approaches μ , and how quickly the *empirical average*

$$\mu_t g := \frac{1}{t} \int_0^t g(X_s) \, ds$$

approaches $\mu g := \int g \, d\mu$, the equilibrium average.

If the Markov process is started from stationarity with $X_0 \sim \mu$, the convergence of empirical averages is naturally measured by the *empirical variance*

$$\text{Var}_\mu(\mu_t g) := \mathbb{E}_\mu |\mu_t g - \mu g|^2.$$

In the context of Markov chain Monte Carlo (MCMC) algorithms, the two questions about convergence raised above target, respectively, the burn-in time and the time required for accurate sampling after (approximate) stationarity is reached; see e.g. [HJ01] [RR04].

In the present paper, we study the same basic convergence questions but depart from the traditional perspectives in some respects. There is already substantial research on nonreversible Markov chains and processes—see the review in section 1.9 of [Cha25] and section 1.3 below. The interest in such processes has been mainly driven by their accelerated convergence. While most previous studies focus on specific classes of examples, we follow the work on nonreversible finite-state Markov chains in [Cha25] and consider a universal spectral theory for nonreversible Markov processes on arbitrary state spaces. Even without reversibility, we find that a generalization of the spectral gap that we call the singular-value gap remains useful and informative. The main finding is that a positive singular-value gap gives finite-time, observable-independent bounds on empirical variance and naturally relates to other quantities for Markov process convergence.

1.1.1. *Spectral gap vs. singular-value gap.* In the reversible setting, different notions of convergence are tied to the same spectral quantity. Since $-L$ is nonnegative and self-adjoint on $L^2(\mu)$, its spectrum is contained in $\mathbb{R}_{\geq 0}$. The *spectral gap* λ of L is then defined as

$$\lambda := \inf \sigma(-L \mid_{L_0^2(\mu)}),$$

where σ denotes the spectrum. For a discrete-time finite-state Markov chain with transition matrix P , the generator is given by $L = P - I$, in which case λ is usually characterized as one minus the second-largest eigenvalue of P . When $\lambda > 0$, we have exponential L^2 -contraction of the semigroup: $\|P_t f\| \leq e^{-\lambda t} \|f\|$ for all centered observables f . This leads to the Poincaré inequality $\|f\|^2 \leq \lambda^{-1} \langle f, -Lf \rangle$, where $\langle -Lf, f \rangle$ is the Dirichlet form evaluated on f and is often written as $\mathcal{E}(f, f)$. In fact, the spectral gap can be characterized as (the inverse of) the optimal constant in the Poincaré inequality:

$$\lambda = \inf_{f \in \text{Dom}(L) \cap L_0^2(\mu) \setminus \{0\}} \frac{\langle -Lf, f \rangle}{\|f\|^2}. \quad (1)$$

Although less emphasized in the continuous-time literature, it is a quick consequence of L^2 -contraction and the standard identity (16) that when the process is reversible,

$$\text{Var}_\mu(\mu_t g) \leq \frac{2}{\lambda t} \|g\|^2 \quad (2)$$

for all $g \in L_0^2(\mu)$. For Markov chains on finite state space, the time- n empirical average is

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

and the corresponding empirical variance bound holds exactly as in (2) for a reversible chain, with continuous t replaced by discrete n ; see Lemma 12.22 in [LP17]. This perspective of the spectral gap as a canonical convergence quantity is central in the classical theory of reversible Markov chains and Markov diffusion processes; see [LP17] [BGL14] and references therein.

For nonreversible processes, the situation is more delicate, as the generator is no longer self-adjoint. There are multiple ways to retrieve a spectral theory: one may still consider the spectral gap γ_s of the symmetrized process whose generator is given by

$$L_s := \frac{L + L^*}{2}; \quad (3)$$

see the more precise definition of γ_s in (32) below. Alternatively, one may also consider the real-part spectral gap of L or resolvent and pseudospectral quantities. There is no longer a single canonical spectral object.

For nonreversible finite-state Markov chains, it is demonstrated in [Cha25] that empirical variance of observables can be uniformly bounded by the singular-value gap γ , defined as the second-smallest singular value of $I - P$, where P is the transition matrix. Note that $L = P - I$ is the generator in discrete time, and we may equivalently write

$$\gamma := \inf_{f \in L_0^2(\mu) \setminus \{0\}} \frac{\|(P - I)f\|}{\|f\|}. \quad (4)$$

With the formulation in (4), it is natural to consider the continuous-time analog

$$\gamma := \inf_{f \in \text{Dom}(L) \cap L_0^2(\mu) \setminus \{0\}} \frac{\|Lf\|}{\|f\|}. \quad (5)$$

Indeed, this *singular-value gap* has also been studied in the context of continuous-time Markov processes, most recently in [EL24][HL25]. The terminology “singular-value gap” is literal. Let $L_0 := L|_{L_0^2(\mu)}$, with $\text{Dom}(L_0) = \text{Dom}(L) \cap L_0^2(\mu)$. Consider the nonnegative self-adjoint operator $|L_0| := (L_0^* L_0)^{1/2}$. By the polar decomposition for closed densely defined operators, $\|L_0 f\| = \||L_0|f\|$ for $f \in \text{Dom}(L_0)$. Therefore,

$$\gamma = \inf_{f \in \text{Dom}(L_0), \|f\|=1} \|L_0 f\| = \inf \sigma(|L_0|).$$

Equivalently,

$$\gamma^2 = \inf \sigma(L_0^* L_0).$$

Thus, in a finite state space, γ is exactly the second-smallest singular value of the matrix $-L$. When L is normal, by the spectral theorem, we also have

$$\gamma = \text{dist}(0, \sigma(L_0)),$$

but this interpretation can fail without normality.

As we shall show, among various inequivalent metrics of convergence for a nonreversible Markov process, the singular-value gap γ defined in (5) specifically characterizes the convergence for empirical averages. This can be intuitively understood through the following simple example, which distinguishes γ from the spectral gap γ_s of the symmetric part of the generator.

Example 1. Consider Brownian motion with constant drift on the unit circle $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$:

$$dX_t = \alpha dt + \sqrt{2} dB_t \quad \text{mod } 2\pi, \quad \text{with generator } L_\alpha = \partial_\theta^2 + \alpha \partial_\theta.$$

The invariant measure μ is normalized Lebesgue measure on $\mathbb{T}$. An orthonormal basis for $L^2(\mu)$ is given by the Fourier modes $e_k(\theta) = e^{ik\theta}$ for $k \in \mathbb{Z}$, where $(e_k)_{k \neq 0}$ form a basis for $L_0^2(\mu)$. Each e_k is an eigenfunction for $-L_\alpha$ with eigenvalue $k^2 - i\alpha k$, so we get

$$\gamma = \inf_{k \neq 0} |k^2 - i\alpha k| = \sqrt{1 + \alpha^2}.$$

On the other hand, the symmetric part of L_α is $L_s = \partial_\theta^2$, and the symmetrized gap is

$$\gamma_s = \inf_{k \neq 0} k^2 = 1.$$

One may also check that $P_t e_k = e^{-(k^2 - i\alpha k)t} e_k$, hence $\|P_t\|_{L_0^2 \rightarrow L_0^2} = e^{-t}$. The L^2 -contraction rate is correctly captured by γ_s . In fact, it is in general true that positive γ_s implies exponential contraction of P_t ; see (40) below. However, one may check $\lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t e_k) = \frac{2}{k^2 + \alpha^2}$, so

$$\sup_{g \in L_0^2(\mu): \|g\|=1} \lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t g) = \frac{2}{1 + \alpha^2} \asymp \frac{1}{\gamma^2}.$$

The convergence speed of empirical averages is captured by γ but missed by γ_s .

The observations above are consistent with the intuition that a drift of strength α should accelerate the convergence of empirical averages, but it merely translates the law of X_t by αt and does not help with its convergence to the uniform distribution.

1.1.2. *Asymptotic variance vs. empirical variance.* Since $\text{Var}_\mu(\mu_t g)$ often decays as t^{-1} , it is of interest to study the *asymptotic variance*, which is naturally connected to functional CLT results. For $g \in L_0^2(\mu)$, the asymptotic variance of g , when it exists, is defined by

$$\sigma_g^2 := \lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t g). \quad (6)$$

We will call $\sigma_g \geq 0$ the *asymptotic standard deviation* of g governing long-time fluctuation:

$$\sqrt{\text{Var}_\mu(\mu_t g)} \sim \frac{\sigma_g}{\sqrt{t}}.$$

A standard way to analyze σ_g^2 is through the Poisson equation. If we can find $f \in \text{Dom}(L)$ that solves $-Lf = g$, then

$$\sigma_g^2 = 2 \text{Re}\langle g, f \rangle. \quad (7)$$

See Lemma 5 for a short derivation of (7). When we can solve the Poisson equation for g and the process satisfies a few other technical assumptions, Dynkin's formula and central limit theorem for martingales imply that $\sqrt{t} \mu_t g$ converges to a Gaussian with mean zero and variance σ_g^2 . This is the classical CLT result of [Bha82]. A substantial theory of functional CLT, notably initiated by [KV86], establishes analogous results without requiring a solution to the Poisson equation. See Chapter 2 of [KLO12] and the references therein.

For the purposes of this paper, given a positive γ , we are in the simpler regime that the Poisson equation is solvable. In fact, L^{-1} exists on $L_0^2(\mu)$ with $\|L^{-1}\|_{L_0^2 \rightarrow L_0^2} \leq \gamma^{-1}$, thanks to Lemma 2. Therefore, the Poisson equation is solvable for all functions in $L_0^2(\mu)$. This yields a CLT under extra technical assumptions; for instance, when the process is progressively measurable, by Theorem 2.1 in [Bha82]. Even if we only know that $\gamma > 0$, we have a simple bound for the asymptotic variance:

$$\sigma_g^2 = 2 \text{Re}\langle g, -L^{-1}g \rangle \leq \frac{2}{\gamma} \|g\|^2. \quad (8)$$

In the above bound, the exact value of σ_g^2 may be much smaller and dependent on the particular observable g . The singular-value gap studied in this paper provides observable-class-level bounds on finite-time empirical variance and asymptotic variance. The asymptotic variance is usually studied in the literature to obtain sharp long-time behavior for specific observables. Knowledge of asymptotic variance helps refine bounds on finite-time empirical variance: see Theorem 3 below.

1.2. Main results. We shall state our main results in a subspace-dependent form. Let $V \subset L_0^2(\mu)$ be a subspace. Consider the singular-value gap of L on V given by

$$\gamma_V := \inf_{f \in \text{Dom}(L) \cap V \setminus \{0\}} \frac{\|Lf\|}{\|f\|}. \quad (9)$$

Let $U := L(V \cap \text{Dom}(L))$. Note that since μ is invariant, $\int Lf d\mu = 0$ for $f \in \text{Dom}(L) \cap L_0^2(\mu)$; hence $U \subset L_0^2(\mu)$. We quantify the worst-case convergence of empirical averages of observables in U by the quantity

$$\Delta_t^U := \sup_{g \in U: \|g\|=1} \sqrt{\text{Var}_\mu(\mu_t g)} = \sup_{g \in U + \mathbb{C}\mathbf{1}: \|g - \mu g\|=1} \sqrt{\text{Var}_\mu(\mu_t g)} \quad (10)$$

We have the following upper and lower bounds on Δ_t^U in terms of γ_V :

Theorem 1. *In the above setting, assume $\gamma_V > 0$. Then*

$$(i) \quad \Delta_t^U \leq \frac{2}{\sqrt{t\gamma_V}}$$

$$(ii) \quad \sup_{s \in [t, 2t]} \Delta_s^{U+V} \geq \frac{1}{\frac{3}{2}t\gamma_V + 3}$$

With the above theorem and the key input of Lemma 2, we obtain control of the empirical standard deviation of all observables in $L^2(\mu)$, quantified by

$$\Delta_t := \sup_{g \in L_0^2(\mu): \|g\|=1} \sqrt{\text{Var}_\mu(\mu_t g)} = \sup_{g \in L^2(\mu): \|g - \mu g\|=1} \sqrt{\text{Var}_\mu(\mu_t g)} \quad (11)$$

Corollary 1. *Assume the global singular-value gap γ defined in (5) satisfies $\gamma > 0$. Then*

$$(i) \quad \Delta_t \leq \frac{2}{\sqrt{\gamma t}}$$

$$(ii) \quad \sup_{s \in [t, 2t]} \Delta_s \geq \frac{1}{\frac{3}{2}\gamma t + 3}$$

Note that for a pure rotation on the unit circle, for instance, the empirical average is exactly the equilibrium average at times when an integer number of rotations are performed. Therefore, for a general nonreversible dynamics, it seems infeasible to obtain lower bounds for empirical variance without taking the worst case over a time interval.

One might also wonder whether the bounds above are optimal. We discuss simple examples in section 2.3 that show that the upper and lower bounds in Corollary 1 are generally sharp up to constants. However, Theorem 3 below gives the following refinement in terms of the asymptotic variance:

$$\frac{\sigma_g^2}{t} - \frac{4}{\gamma^2 t^2} \|g\|^2 \leq \text{Var}_\mu(\mu_t g) \leq \frac{\sigma_g^2}{t} + \frac{4}{\gamma^2 t^2} \|g\|^2 \quad (12)$$

In particular, when $\sigma_g^2 > 0$ for observables g under consideration, the standard deviation of empirical averages of g decays as $t^{-1/2}$, the expected CLT scale. For certain dynamics and observables, the empirical standard deviation can decay faster on the order of t^{-1} ; we discuss a simple instance in Example 4. The bound (12) shows that such a faster convergence happens exactly when the asymptotic variance vanishes, but we do not explore the separation between the $t^{-1/2}$ and t^{-1} regimes in more detail.

1.2.1. Outline. We contextualize our perspective within the existing literature in section 1.3. In section 2, we prove the local Theorem 1 and deduce the global Corollary 1 from Lemma 2, which gives a bounded inverse generator when $\gamma > 0$. Using this lemma, we also extend Chatterjee's original theorem for finite-state Markov chains to arbitrary state spaces, confirming that the singular-value gap γ is a universal quantity for convergence. We then discuss the sharpness of our variance bounds and prove the sharpened version (12) in terms of asymptotic variance.

In section 3, we study in detail the example of a two-dimensional OU process with rotational drift. This example is inspired by [DLP16] and sits within a substantial literature on accelerating the convergence of diffusion processes by adding an irreversible drift. The main point of this example is that irreversibility may accelerate the convergence of empirical averages only for certain subspaces of observables.

We explore the connections between γ and some other common quantities related to mixing in section 4. In section 4.1, we bound γ in terms of γ_s , the spectral gap of the symmetrized process. Theorem 4 shows that γ^{-1} provides a lower bound for the mixing time in total-variation distance, and Proposition 5 gives an upper bound on the mixing time in the restricted setting of jump processes on finite state space. We then connect the singular-value gap to the Cheeger constant, also known as the bottleneck ratio. Theorem 5 extends the previous results of [LS88][Cha25] to nonreversible jump processes.

Our theory can be applied under minimal assumptions, but we only survey two more examples in section 5. We consider random walks on binary de Bruijn graphs to show that, as γ^{-1} lower bounds the total-variation mixing time by Theorem 4, the singular-value gap could more sharply record mixing properties compared to some other spectral quantities. We then observe that a compact resolvent criterion in Proposition 6 is enough to show that $\gamma > 0$ for certain Markov diffusion generators, without explicit knowledge of the value of γ . We study a specific hypoelliptic generator on the 2-torus and show that $\gamma > 0$, hence empirical averages converge with fluctuations of order at most $t^{-1/2}$. This example provides an interesting bridge between abstract PDE techniques and concrete sampling variance estimates. It would be interesting to find more substantial applications of the singular-value gap beyond the examples given in this paper.

1.3. Related literature. As mentioned above, the closest predecessor to the present work is the theory of nonreversible spectral gaps developed in [Cha25] for nonreversible finite-state Markov chains. Chatterjee shows that for a chain with transition matrix P , the reciprocal of the gap γ defined as the second-smallest singular value of $I - P$ characterizes the time scale for the convergence of empirical averages. He also relates this quantity to mixing time, Cheeger constants, and path arguments. Our work follows this generator-singular-value viewpoint but extends it to continuous time, general state space, and localized observable classes.

An earlier precursor is [Din98], which obtains deviation bounds for the empirical averages of nonreversible chains in terms of the singular-value gap of the discrete generator. More recently, [EL24] discusses the singular-value gap for nonreversible diffusion lifts, although the main relaxation notion there is an L^2 -contraction time of the semigroup rather than empirical-average variance. These generator singular values should be distinguished from other nonreversible spectral quantities. Notably, for discrete-time chains, Fill’s multiplicative reversibilization [Fil91] and Paulin’s pseudo-spectral gap [Pau18] use singular values of transition operators, or of products such as $(P^*)^k P^k$, to control convergence and concentration. [KM11] emphasize yet another perspective: for general state-space nonreversible chains, spectral theory for geometric ergodicity is often better formulated in weighted L^∞ -spaces. In the present paper, we keep our focus on $L^2(\mu)$.

There is a large literature showing that nonreversibility can improve convergence, hence facilitate sampling. In finite state space, lifting is a central mechanism. [CLP99] shows that lifting can speed up mixing and that square-root-type speedups are optimal in broad settings. [DHN00] analyzes a concrete lifted nonreversible sampler. Analogous bounds for lifting constructions on continuous state space are developed in [RS18].

For diffusions, [HHS93][HHS05] show that adding a divergence-free drift preserving a target distribution can accelerate convergence. [LNP13] studies optimal nonreversible linear drifts, emphasizing both exponential rates and prefactors. [HNW15] gives general comparison results showing that antisymmetric perturbations reduce asymptotic variance for diffusions. There are also more detailed investigations of variance reduction: [RS15] uses large deviations of the empirical measure to quantify variance reduction for irreversible Langevin samplers, while [DLP16] analyzes variance reduction via the Poisson equation and calculations for linear and quadratic observables. Our results are complementary: rather than considering irreversible perturbations of a certain form or reducing observable-specific asymptotic variance, we ask what can be concluded from a positive singular-value gap of the generator, globally or on an observable subspace. The connections of our perspective with asymptotic variance and central-limit theorems are discussed in section 1.1.2 above.

There is also substantial literature on finite-time deviation bounds for empirical averages. In the reversible setting, variance bounds for stationary empirical averages follow from the spectral gap, as recorded in (2). [Lez01] proves Chernoff and Berry–Esséen inequalities for empirical means of general ergodic Markov processes with a spectral gap. [Pau18] obtains concentration, variance and Bernstein-type bounds for possibly nonreversible Markov chains through the pseudo-spectral gap. [JO10] derives nonasymptotic empirical-average error estimates under a positive-curvature assumption. More recently, Bernstein-type inequalities are obtained in [HL25], using the singular-value gap (4), which the authors call the iterated Poincaré gap. In continuous time, [HL25] uses a slightly different characterization of the singular-value gap (5) and proves concentration inequalities for empirical averages of bounded observables, allowing nonstationary starts.

In contrast to some aforementioned studies with MCMC sampling in mind, we work primarily under stationarity. In applications where the chain is only approximately stationary, one can combine burn-in, warm-start, or concentration estimates with the stationary bounds studied here; developing this systematically is an interesting direction beyond the scope of the present paper.

2. SINGULAR-VALUE GAP CONTROLS EMPIRICAL VARIANCE

We will study the empirical average $\mu_t g$ through the Cesàro mean

$$A_t g := \frac{1}{t} \int_0^t P_s g \, ds. \quad (13)$$

Note that $A_t g$ is a function on the state space S , whereas $\mu_t g$ is a $\mathbb{C}$ -valued random variable. They are related by

$$\mathbb{E}_x[\mu_t g] = A_t g(x)$$

Consider a subspace $U \subset L_0^2(\mu)$. Analogous to Δ_t^U, Δ_t in (10) (11), we define

$$\begin{aligned}\delta_t^U &:= \sup_{g \in U: \|g\|=1} \|A_t g\| = \sup_{g \in U + \mathbb{C}\mathbf{1}: \|g - \mu g\|=1} \|A_t g - \mu g\|, \\ \delta_t &:= \sup_{g \in L_0^2(\mu): \|g\|=1} \|A_t g\| = \sup_{g \in L^2(\mu): \|g - \mu g\|=1} \|A_t g - \mu g\|.\end{aligned}$$

We record a simple lemma for future use:

Lemma 1 (Well-known stationary covariance identity). *For $u \geq s$ and $g, h \in L^2(\mu)$,*

$$\mathbb{E}_\mu[\overline{g(X_s)} h(X_u)] = \langle g, P_{u-s} h \rangle.$$

Proof. By the Markov property,

$$\mathbb{E}_\mu[\overline{g(X_s)} h(X_u)] = \mathbb{E}_\mu[\overline{g(X_s)} \mathbb{E}_\mu[h(X_u) \mid \mathcal{F}_s]] = \mathbb{E}_\mu[\overline{g(X_s)} (P_{u-s} h)(X_s)].$$

Since the process is stationary, $X_s \sim \mu$, so

$$\mathbb{E}_\mu[\overline{g(X_s)} h(X_u)] = \int \overline{g(x)} (P_{u-s} h)(x) \mu(dx) = \langle g, P_{u-s} h \rangle.$$

□

2.1. Local singular-value gap.

2.1.1. *Upper bound in Theorem 1.* The following proposition is inspired by Lemma 3.1 in [Cha25].

Proposition 1 (Cesàro-average upper bound). *Let $V \subset L_0^2(\mu)$ be a subspace such that $U = L(V \cap \text{Dom}(L)) \subset L_0^2(\mu)$. Assume that the local gap defined in (9) satisfies $\gamma_V > 0$. Then*

$$\delta_t^U \leq \frac{2}{t\gamma_V}.$$

Proof. Consider a function $g \in U$. Choose $f \in V \cap \text{Dom}(L)$ such that $-Lf = g$. Using the semigroup identity $\int_0^t P_s Lf ds = P_t f - f$ (see, for instance, Lemma II.1.3 (iv) in [EN00]), we get $\int_0^t P_s g ds = -\int_0^t P_s Lf ds = f - P_t f$. Dividing by t yields

$$A_t g = \frac{1}{t}(f - P_t f). \tag{14}$$

Now P_t is an $L^2(\mu)$ -contraction, so

$$\|A_t g\| \leq \frac{1}{t}(\|f\| + \|P_t f\|) \leq \frac{2}{t}\|f\|.$$

By the definition of γ_V , $\|g\| = \|Lf\| \geq \gamma_V \|f\|$, hence

$$\|f\| \leq \frac{1}{\gamma_V} \|g\|.$$

Combining the two inequalities above, we get

$$\|A_t g\| \leq \frac{2}{t\gamma_V} \|g\|.$$

□

With the above proposition and Lemma 1, we are ready to prove the upper bound on the empirical variance of observables in $U \subset L_0^2(\mu)$. The identities (15) (16) are standard in the literature on the ergodic average of Markov processes.

Proof of upper bound in Theorem 1. Let $g \in U$. With $\mu_t g = \frac{1}{t} \int_0^t g(X_s) ds$, we have

$$\begin{aligned} \text{Var}_\mu(\mu_t g) &= \frac{1}{t^2} \mathbb{E}_\mu \left| \int_0^t g(X_s) ds \right|^2 \\ &= \frac{1}{t^2} \int_0^t \int_0^t \mathbb{E}_\mu [\overline{g(X_s)} g(X_u)] ds du \\ &= \frac{2}{t^2} \text{Re} \int_0^t \int_0^u \mathbb{E}_\mu [\overline{g(X_s)} g(X_u)] ds du \\ &= \frac{2}{t^2} \text{Re} \int_0^t \int_0^u \langle g, P_{u-s} g \rangle ds du \\ &= \frac{2}{t^2} \text{Re} \int_0^t \int_0^u \langle g, P_r g \rangle dr du \\ &= \frac{2}{t^2} \text{Re} \int_0^t (t-r) \langle g, P_r g \rangle dr. \end{aligned} \tag{15}$$

$$= \frac{2}{t^2} \text{Re} \int_0^t (t-r) \langle g, P_r g \rangle dr. \tag{16}$$

For $0 \leq u \leq t$, define

$$h_u := \int_0^u P_r g dr = u A_u g.$$

Then

$$\text{Var}_\mu(\mu_t g) = \frac{2}{t^2} \int_0^t \text{Re} \langle g, h_u \rangle du.$$

By Cauchy–Schwarz and Proposition 1,

$$\text{Re} \langle g, h_u \rangle \leq |\langle g, h_u \rangle| \leq \|g\| \|h_u\| = u \|g\| \|A_u g\| \leq \frac{2}{\gamma_V} \|g\|^2.$$

Hence

$$\text{Var}_\mu(\mu_t g) \leq \frac{2}{t^2} \int_0^t \frac{2}{\gamma_V} \|g\|^2 du = \frac{4}{t\gamma_V} \|g\|^2. \tag{17}$$

□

Remark 1 (Alternative sufficient condition). *If we start with a subspace $U \subset L_0^2(\mu)$ with a locally bounded Poisson equation solver $R_U : U \rightarrow \text{Dom}(L)$ such that for all $g \in U$*

$$-LR_U g = g, \quad \|R_U g\| \leq \kappa_U \|g\|,$$

then the above proof and conclusion still hold with κ_U^{-1} playing the role of γ_V .

Remark 2 (Constant improvement to empirical variance upper bound). Section 3 in [EL24] considers the global singular-value gap γ in connection with $L_0^2(\mu)$ -relaxation of nonreversible lifts. It is also claimed there that for $g \in L_0^2(\mu)$,

$$\text{Var}_\mu(\mu_t g) \leq \frac{2}{\gamma t} \|g\|^2.$$

We are not able to find an explicit proof for this claimed upper bound on the finite-time variance of the empirical average. This bound as written does not follow from the argument used here, but we do have a factor of 2 improvement to (17) for large time, thanks to Theorem 3:

$$\text{Var}_\mu(\mu_t g) \leq \left(\frac{2}{\gamma t} + \frac{4}{\gamma^2 t^2} \right) \|g\|^2.$$

2.1.2. *Lower bound in Theorem 1.* Similarly to the proof of the upper bound, we control Δ_t^{U+V} through δ_t^{U+V} . It is perhaps unnecessary that the lower bound involves observables in V as well as U ; a more refined understanding of the lower bound would be interesting.

Proposition 2. For every $t > 0$,

$$\sup_{s \in [t, 2t]} \delta_s^{U+V} \geq \frac{1}{\frac{3}{2}t\gamma_V + 3}.$$

Proof. Fix $t > 0$, and set

$$M_t := \sup_{s \in [t, 2t]} \delta_s^{U+V}.$$

Let $\varepsilon > 0$. By the definition of γ_V , there exists $f \in V$ such that

$$\|f\| = 1, \quad \|Lf\| \leq \gamma_V + \varepsilon.$$

Set

$$g := -Lf \in U.$$

Now define

$$u_t := f - \frac{1}{t} \int_t^{2t} P_s f \, ds, \quad v_t := \frac{1}{t} \int_t^{2t} P_s f \, ds.$$

Then $f = u_t + v_t$, so

$$1 = \|f\| \leq \|u_t\| + \|v_t\|.$$

By (14),

$$u_t = \frac{1}{t} \int_t^{2t} (f - P_s f) \, ds = \frac{1}{t} \int_t^{2t} s A_s g \, ds.$$

Taking norms, we get

$$\begin{aligned} \|u_t\| &\leq \frac{1}{t} \int_t^{2t} s \|A_s g\| \, ds \\ &\leq \frac{1}{t} \int_t^{2t} s \delta_s^U \|g\| \, ds \\ &\leq \frac{M_t \|g\|}{t} \int_t^{2t} s \, ds \\ &\leq \frac{3}{2} t M_t (\gamma_V + \varepsilon). \end{aligned}$$

Next,

$$v_t = \frac{1}{t} \int_t^{2t} P_s f \, ds = 2A_{2t}f - A_t f,$$

hence

$$\|v_t\| \leq 2\|A_{2t}f\| + \|A_t f\| \leq (2\delta_{2t}^V + \delta_t^V)\|f\| \leq 3M_t.$$

Combining the estimates,

$$1 \leq \|u_t\| + \|v_t\| \leq \left(\frac{3}{2}t(\gamma_V + \varepsilon) + 3 \right) M_t.$$

Letting $\varepsilon \downarrow 0$, we obtain

$$\sup_{t \leq s \leq 2t} \delta_s^{U+V} = M_t \geq \frac{1}{\frac{3}{2}t\gamma_V + 3}.$$

□

Proof of lower bound in Theorem 1. Recall that for $h \in L_0^2(\mu)$,

$$A_t h(x) = \mathbb{E}_x[\mu_t h].$$

By Jensen's inequality,

$$\|A_t h\| = \left(\int |\mathbb{E}_x[\mu_t h]|^2 d\mu(x) \right)^{1/2} \leq (\mathbb{E}_\mu |\mu_t h|^2)^{1/2} = \sqrt{\text{Var}_\mu(\mu_t h)}. \quad (18)$$

Taking the supremum over $h \in U + V$, $\|h\| = 1$ gives

$$\delta_t^{U+V} \leq \Delta_t^{U+V}.$$

Therefore, by Proposition 2,

$$\sup_{s \in [t, 2t]} \Delta_s^{U+V} \geq \frac{1}{\frac{3}{2}t\gamma_V + 3}.$$

□

As in [Cha25], the dyadic lower bound also implies a uniform lower bound at small times, showing that empirical averages cannot converge in a time much shorter than γ^{-1} . We shall state the global version here for simplicity.

Corollary 2 (A small-time lower bound). *For $t \leq 1/(3\gamma)$, we have*

$$\Delta_t \geq \frac{1}{36}$$

Proof. Define

$$F(t) := t\Delta_t.$$

Then F is subadditive. Indeed, for $s, t > 0$, stationarity gives

$$(s+t)\mu_{s+t}g = t \cdot \frac{1}{t} \int_0^t g(X_r) dr + s \cdot \frac{1}{s} \int_t^{t+s} g(X_r) dr.$$

Taking the $L^2(\mathbb{P}_\mu)$ -norms, using the triangular inequality, and then taking the supremum over $\|g\| = 1$, we obtain

$$(s+t)\Delta_{s+t} \leq t\Delta_t + s\Delta_s.$$

Thus, $F(s+t) \leq F(s) + F(t)$. We also have the trivial bound $\Delta_t \leq 1$. Indeed, for $\|g\| = 1$, Jensen's inequality and stationarity give

$$\mathbb{E}_\mu |\mu_t g|^2 = \mathbb{E}_\mu \left| \frac{1}{t} \int_0^t g(X_r) dr \right|^2 \leq \frac{1}{t} \int_0^t \mathbb{E}_\mu |g(X_r)|^2 dr = \|g\|^2 = 1.$$

Now, suppose $0 < t \leq 1/(3\gamma)$. Applying the lower bound in Corollary 1 at time scale $12t$, there exists $s \in [12t, 24t]$ such that

$$\Delta_s \geq \frac{1}{\frac{3}{2}\gamma(12t) + 3} = \frac{1}{18\gamma t + 3} \geq \frac{1}{9}.$$

Write

$$s = nt + r, \quad n = \lfloor s/t \rfloor, \quad 0 \leq r < t.$$

Since $s \geq 12t$, we have $r/s \leq 1/12$. By subadditivity,

$$s\Delta_s = F(s) \leq nF(t) + F(r) \leq nt\Delta_t + r,$$

where the last step uses $\Delta_r \leq 1$. Dividing by $s = nt + r$ gives

$$\Delta_s \leq \Delta_t + \frac{r}{s} \leq \Delta_t + \frac{1}{12}.$$

Therefore, for $0 < t \leq 1/(3\gamma)$,

$$\Delta_t \geq \frac{1}{9} - \frac{1}{12} = \frac{1}{36}.$$

□

2.2. Global singular-value gap. If we want a universal understanding of all observables and take $U = L_0^2(\mu)$, it is a priori unclear from the local statement of Theorem 1 that we should be able to find $f \in L_0^2(\mu)$ that solves the Poisson equation $-Lf = g$ for every $g \in L_0^2(\mu)$. The key observation in Lemma 2 below is that if L possesses a global singular-value gap $\gamma > 0$, then L is in fact invertible on $L_0^2(\mu)$. In particular, we get a bounded Poisson solver that facilitates our proof strategies. Applying this observation to discrete-time Markov chains, we note that Chatterjee's main theorem in [Cha25] can be extended to general state space.

Proof of Corollary 1. By Lemma 2, when $\gamma > 0$, L is invertible on $L_0^2(\mu)$. Then the corollary follows from Theorem 1 by taking $V = U = L_0^2(\mu)$. □

2.2.1. *Positive gap implies bounded Poisson solver.* We state the following lemma in a general form, so that it can be applied to discrete-time Markov chains and invariant subspaces K . For instance, K can be an eigensubspace when the generator L has discrete spectrum.

Lemma 2. *Let (S, μ) be a probability space. Let L be the generator on $L^2(\mu)$, either of a strongly continuous Markov semigroup $(P_t)_{t \geq 0}$, or of the form $L = P - I$ for a discrete-time Markov chain with Markov transition kernel P . Let $K \subset L^2_0(\mu)$ be a closed subspace such that $P_t(K) \subset K$ for all $t \geq 0$ in continuous time, or $P(K) \subset K$ in discrete time. Let L_K denote the generator restricted to K . Suppose that L has a singular-value gap on K :*

$$\gamma_K := \inf_{f \in \text{Dom}(L_K) \setminus \{0\}} \frac{\|L_K f\|}{\|f\|} > 0. \quad (19)$$

Then L_K is invertible with

$$\|L_K^{-1}\|_{K \rightarrow K} \leq \frac{1}{\gamma_K}.$$

Proof. Since L_K is the generator of a contraction semigroup, the resolvent set of L_K contains the positive half-plane. Explicitly, in continuous time, the resolvent R_λ of L_K at $\lambda > 0$ is given by

$$R_\lambda g := (\lambda I - L_K)^{-1} g = \int_0^\infty e^{-\lambda s} P_s g \, ds, \quad g \in K;$$

see Theorem II.1.10 in [EN00]. In discrete time, for $\lambda > 0$,

$$R_\lambda := (\lambda I - L_K)^{-1} = ((1 + \lambda)I - P_K)^{-1},$$

where P_K is the restriction of the Markov kernel P on K . Since $\|\frac{1}{1+\lambda}P_K\| \leq \frac{1}{1+\lambda} < 1$, the Neumann series gives

$$R_\lambda = \frac{1}{1+\lambda} \sum_{n=0}^{\infty} \left(\frac{P_K}{1+\lambda} \right)^n.$$

For $0 < \lambda < \gamma_K$ and $f \in \text{Dom}(L_K)$,

$$\|(\lambda I - L_K)f\| \geq \|L_K f\| - \lambda \|f\| \geq (\gamma_K - \lambda)\|f\|.$$

Therefore, $\|R_\lambda\| \leq \frac{1}{\gamma_K - \lambda}$. For $0 < \lambda, \mu < \gamma_K$, the resolvent identity yields

$$R_\lambda - R_\mu = (\mu - \lambda)R_\lambda R_\mu.$$

Thus, $(R_\lambda)_{\lambda \downarrow 0}$ is Cauchy in operator norm. Let

$$R_K := \lim_{\lambda \downarrow 0} R_\lambda \quad \text{in } \|\cdot\|_{K \rightarrow K}.$$

We claim that $R_K = -L_K^{-1}$. Consider $g \in K$. Since for $\lambda > 0$

$$(\lambda I - L_K)R_\lambda = I,$$

we have

$$-L_K R_\lambda g = g - \lambda R_\lambda g \rightarrow g \quad \text{as } \lambda \downarrow 0.$$

Since L_K is closed, it follows that $R_K g \in \text{Dom}(L_K)$ and

$$-L_K R_K g = g.$$

Similarly, we deduce $-R_K L_K f = f$ for any $f \in \text{Dom}(L_K)$, hence

$$R_K = -L_K^{-1}.$$

Finally, since $\|L_K f\| \geq \gamma_K \|f\|$ for $f \in \text{Dom}(L_K)$, we get $\|L_K^{-1}\| \leq \gamma_K^{-1}$. $\square$

Remark 3. In continuous time, the above proof could be streamlined by appealing to well-known results in semigroup theory: for example, Corollary II.3.6 and Proposition IV.1.3 (iii) in [EN00]. We choose to present explicit details to treat discrete-time Markov chains on the same footing.

Remark 4 (Relation to the potential operator). Let $K = L_0^2(\mu)$ and denote by L_0 the restricted generator. When the Markov process is reversible, a positive spectral gap γ implies exponential decay of $\|P_t g\|$ for $g \in L_0^2(\mu)$, and one may write $-L_0^{-1}$ as the potential operator G defined by

$$Gg := \int_0^\infty P_r g \, dr. \quad (20)$$

In the general nonreversible case, it is possible for γ to be positive, therefore $-L_0^{-1}$ exists, while Gg is not well-defined for some $g \in L_0^2(\mu)$. See Example 2 below. Similarly, a positive γ for a discrete-time chain does not guarantee that P is strictly contractive on $L_0^2(\mu)$, so one cannot always hope to write $-L_0^{-1}$ as the potential operator G defined by

$$Gg := \sum_{k=0}^\infty P^k g.$$

In both continuous time and discrete time, if Gg converges strongly for all $g \in L_0^2(\mu)$, then L_0 is invertible, and $Gg = -L_0^{-1}g$. In continuous time, this can be ensured if $\|P_t\|_{L_0^2 \rightarrow L_0^2} \leq Ce^{-at}$ for some $C, a > 0$. In particular, P_t becomes exponentially contractive if we perturb the generator so that its symmetric part has a positive spectral gap $\gamma_s > 0$; see (40) below.

In discrete time, the corresponding sufficient condition is $\|P\|_{L_0^2 \rightarrow L_0^2} < 1$. For example, if a Markov chain with transition operator P has a positive singular-value gap $\gamma > 0$, then adding laziness makes the transition operator strictly contractive on $L_0^2(\mu)$. Indeed, for $\tilde{P} = \alpha I + (1-\alpha)P$ with $\alpha \in (0, 1)$, one may check that

$$\|\tilde{P}\|_{L_0^2 \rightarrow L_0^2} \leq \sqrt{1 - \alpha(1-\alpha)\gamma^2} < 1,$$

so that $\sum_{k=0}^\infty \tilde{P}^k$ converges in operator norm to the inverse of $I - \tilde{P}$.

Example 2 (Pure rotation on the circle). As a simple example, consider the unit circle $\mathbb{T}$ with normalized Lebesgue measure μ , where the dynamics is given by deterministic rotation: $P_t g(\theta) := g(\theta + t)$, so that $L = \partial_\theta$. The space $L_0^2(\mu)$ is spanned by the orthonormal Fourier modes $e_k(\theta) = e^{ik\theta}$ for $k \neq 0$. Since $Le_k = ik e_k$, L_0 has singular-value gap $\gamma = 1$ achieved by $g = e_1$, so L_0 is invertible and $-L_0^{-1}g = ig$. On the other hand, the integral defining Gg does not converge:

$$\int_0^t P_s g \, ds = \int_0^t e^{is} \, ds \, g = \frac{e^{it} - 1}{i} g.$$

2.2.2. Variance bounds for Markov chains on general state space. Here we briefly record that Lemma 2 also admits the extension of Theorem 1.2 in [Cha25] to general state space. Let us first clarify the setting. Let $(X_n)_{n=0}^\infty$ be a time-homogeneous discrete time Markov chain on a measurable space S . The Markov transition kernel $P(x, dy)$ acts on $g \in L^2(\mu)$ by $Pg(x) = \int g(y) P(x, dy)$. Assume that the chain is stationary with $X_0 \sim \mu$, where μ is an invariant measure satisfying $\int P(x, dy) \mu(dx) = \mu(dy)$. Use the same notation as above and as in [Cha25]:

$$\mu g := \int g d\mu, \quad \mu_n g := \frac{1}{n} \sum_{i=0}^{n-1} g(X_i)$$

As above, use $\|g\|$ to denote the $L^2(\mu)$ norm of a function g on the state space. The worst-case standard deviation of empirical averages at time n is captured by

$$\Delta_n := \sup_{g \in L_0^2(\mu): \|g\|=1} \sqrt{\text{Var}_\mu(\mu_n g)} = \sup_{g \in L^2(\mu): \|g - \mu g\|=1} \sqrt{\text{Var}_\mu(\mu_n g)}$$

In this case, the generator is $L = P - I$, and we consider the singular-value gap defined in (4). Note that since P is a contraction on $L^2(\mu)$, L is a bounded operator and we do not need to worry about its domain.

We have the exact analog of Theorem 1.2 in [Cha25] in this setting. A similar local statement holds with the same proof, but we only state the global theorem for simplicity.

Theorem 2. Assume $\gamma > 0$. For $n \geq 1$, we have the following upper and lower bounds on Δ_n :

$$(i) \quad \Delta_n \leq \frac{2}{\sqrt{\gamma n}}$$

$$(ii) \quad \max_{n \leq k \leq 2n} \Delta_k \geq \frac{1}{2\gamma n + 3}.$$

Proof. The exact proof in [Cha25] goes through, with the following two modifications. For the upper bound, finite state space only enters the original proof in [Cha25] through a dimension-counting argument to ensure that $(P - I)$ is invertible on $L_0^2(\mu)$ when $\gamma > 0$. This argument can be replaced by Lemma 2.

For the lower bound, finite state space only enters the original proof with the choice of a minimizer f of $\|(P - I)f\|$ with $\|f\| = 1$. This can be resolved by considering near-minimizers, as we did in the proof of Proposition 2. $\square$

2.3. Sharpness of upper and lower bounds. Although they show that $t \gtrsim \gamma^{-1}$ is sufficient and necessary for empirical averages to converge, the upper and lower bounds in Corollary 1 should not be interpreted as a two-sided equivalence for Δ_t . Rather, they capture two different universal obstructions. The upper bound

$$\Delta_t \leq \frac{2}{\sqrt{\gamma t}}$$

is sharp in general. In particular, it has the correct order $t^{-1/2}$ when the observable in question has a finite positive asymptotic variance. On the other hand, the lower bound

$$\sup_{s \in [t, 2t]} \Delta_s \geq \frac{1}{\frac{3}{2}\gamma t + 3}$$

is also sharp as a universal lower bound: without additional assumptions, one cannot replace it by a lower bound of order $t^{-1/2}$. We first illustrate the sharpness of the upper bound with a familiar example.

Example 3 (Stationary iid sampling). *Let (S, μ) be a probability space, and let X_t be the following stationary Markov process. At exponential rate q , the current state is replaced by an independent sample from μ . The generator of this process is*

$$Lf(x) = q \int (f(y) - f(x)) \mu(dy).$$

Thus, on $L_0^2(\mu)$,

$$Lf = -qf, \quad P_t f = e^{-qt} f.$$

Hence the global singular-value gap is

$$\gamma = q.$$

For any $g \in L_0^2(\mu)$, we have $\langle g, P_r g \rangle = e^{-qr} \|g\|^2$. Therefore, using (15),

$$\begin{aligned} \text{Var}_\mu(\mu_t g) &= \frac{2}{t^2} \int_0^t \int_0^u e^{-qr} dr du \|g\|^2 \\ &= \left(\frac{2}{qt} - \frac{2(1 - e^{-qt})}{q^2 t^2} \right) \|g\|^2. \end{aligned}$$

Consequently, as $t \rightarrow \infty$,

$$\Delta_t \sim \sqrt{\frac{2}{\gamma t}}.$$

Thus, the upper bound in Corollary 1 has the correct order in both t and γ , up to constants. The discrete-time analog is the independent-sampling chain $P(x, \cdot) = \mu(\cdot)$. Then $P - I = -I$ on $L_0^2(\mu)$, so $\gamma = 1$, and empirical averages have standard deviation exactly of order $n^{-1/2}$.

The following example shows that the lower bound in Corollary 1 is also sharp. Note that in this case, the lower bound in Theorem 3 below does not imply a $t^{-1/2}$ order for the empirical standard deviation, because the asymptotic variance vanishes.

Example 4 (Pure rotation on the circle, revisited). *Let $S = \mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$, μ be the normalized Lebesgue measure, and consider deterministic rotation*

$$X_t = X_0 + \alpha t \pmod{2\pi}, \quad \alpha \neq 0.$$

The semigroup and the generator are given by

$$P_t g(\theta) = g(\theta + \alpha t), \quad L = \alpha \partial_\theta.$$

For the Fourier modes $e_k(\theta) = e^{ik\theta}$, $k \in \mathbb{Z}$, $Le_k = i\alpha k e_k$, so

$$\gamma = \inf_{k \neq 0} |i\alpha k| = |\alpha|.$$

Since the dynamics is deterministic and μ is invariant, for centered g we have $\mu_t g = A_t g(X_0)$. As $X_0 \sim \mu$, we get

$$\text{Var}_\mu(\mu_t g) = \mathbb{E}_\mu |\mu_t g|^2 = \int |A_t g(x)|^2 \mu(dx) = \|A_t g\|^2. \quad (21)$$

For a Fourier mode,

$$A_t e_k = \frac{1}{t} \int_0^t e^{i\alpha k s} e_k ds = \frac{e^{i\alpha k t} - 1}{i\alpha k t} e_k,$$

hence

$$\Delta_t = \sup_{k \neq 0} \left| \frac{e^{i\alpha k t} - 1}{i\alpha k t} \right| \leq \frac{2}{|\alpha|t} = \frac{2}{\gamma t}.$$

Thus, empirical standard deviation can decay at order $(\gamma t)^{-1}$. In particular, no universal lower bound of order $(\gamma t)^{-1/2}$ can hold under only the assumption of a positive singular-value gap.

Moreover, we can explicitly check that the lower bound of the dyadic supremum in Corollary 1 has the correct order. If $|\alpha|t \geq 2\pi$, then the interval $[t, 2t]$ contains some s for which $e^{i\alpha s} = -1$. For this s ,

$$\Delta_s \geq \left| \frac{e^{i\alpha s} - 1}{i\alpha s} \right| = \frac{2}{|\alpha|s} \geq \frac{1}{|\alpha|t} = \frac{1}{\gamma t}.$$

Thus, for large γt ,

$$\sup_{s \in [t, 2t]} \Delta_s \asymp \frac{1}{\gamma t}.$$

2.4. Sharpened bounds with access to asymptotic variance. The preceding examples show that the singular-value gap alone cannot give a two-sided characterization of Δ_t at the order $t^{-1/2}$. A sharper description is possible when we can locate a set of observables for which the asymptotic variance is computable or can be bounded. In particular, when we have a local gap $\gamma_K > 0$ on a closed invariant subspace K , the asymptotic variance can be uniformly bounded above by γ_K^{-1} . As observed in Remark 2, this provides a constant improvement to the upper bound stated in Theorem 1. More importantly, the lower bound in (24) improves the t -dependence of the lower bound in Theorem 1 when $\sigma_g^2 > 0$ for an observable g in question, bounding the finite-time standard deviation of empirical averages within the CLT scale of $t^{-1/2}$. This leads to an improved observable-class lower bound whenever the worst-case asymptotic variance constant σ_K^2 is positive.

Let $K \subset L_0^2(\mu)$ be a closed invariant subspace to which Lemma 2 applies. Assume $\gamma_K > 0$, so that $L_K^{-1} : K \rightarrow K$ exists and $\|L_K^{-1}\| \leq \gamma_K^{-1}$. For $g \in K$, consider its asymptotic variance σ_g^2 defined in (6). By Lemma 5, we have

$$\sigma_g^2 = 2 \text{Re} \langle g, -L_K^{-1} g \rangle.$$

We also define the worst-case asymptotic variance constant on K by

$$\sigma_K^2 := \sup_{\substack{g \in K \\ \|g\|=1}} \sigma_g^2 = 2 \sup_{\substack{g \in K \\ \|g\|=1}} \text{Re} \langle g, -L_K^{-1} g \rangle. \quad (22)$$

Since $\|L_K^{-1}\| \leq \gamma_K^{-1}$, we have

$$0 \leq \sigma_K^2 \leq \frac{2}{\gamma_K}. \quad (23)$$

Theorem 3 (Finite-time comparison with asymptotic variance). *In the above setting,*

$$\left| \text{Var}_\mu(\mu_t g) - \frac{\sigma_g^2}{t} \right| \leq \frac{4}{\gamma_K^2 t^2} \|g\|^2 \quad (24)$$

for every $g \in K$. Consequently,

$$\frac{\sigma_K^2}{t} - \frac{4}{\gamma_K^2 t^2} \leq (\Delta_t^K)^2 \leq \frac{\sigma_K^2}{t} + \frac{4}{\gamma_K^2 t^2}. \quad (25)$$

Proof. We again use the identity (15):

$$\text{Var}_\mu(\mu_t g) = \frac{2}{t^2} \text{Re} \int_0^t \int_0^u \langle g, P_r g \rangle dr du.$$

Recall that $\int_0^u P_r g ds = f - P_u f$ for $f = -L_K^{-1}g$. Since $-L_K^{-1}$ commutes with the semigroup,

$$\begin{aligned} \int_0^t \int_0^u P_r g dr du &= \int_0^t -L_K^{-1}(I - P_u)g du \\ &= -tL_K^{-1}g + L_K^{-2}(P_t - I)g. \end{aligned}$$

Hence

$$\begin{aligned} \text{Var}_\mu(\mu_t g) &= \frac{2}{t} \text{Re} \langle g, -L_K^{-1}g \rangle + \frac{2}{t^2} \text{Re} \langle g, L_K^{-2}(P_t - I)g \rangle \\ &= \frac{\sigma_g^2}{t} + \frac{2}{t^2} \text{Re} \langle g, L_K^{-2}(P_t - I)g \rangle. \end{aligned}$$

Using $\|L_K^{-1}\| \leq \gamma_K^{-1}$ and $\|P_t\| \leq 1$,

$$|\text{Re} \langle g, L_K^{-2}(P_t - I)g \rangle| \leq |\langle g, L_K^{-2}(I - P_t)g \rangle| \leq \|g\| \|L_K^{-2}\| \|(I - P_t)g\| \leq \frac{2}{\gamma_K^2} \|g\|^2.$$

This proves the first inequality. Taking the supremum over functions $g \in K$ with $\|g\| = 1$ gives the second pair of inequalities. $\square$

We note that when the process is nonreversible, the variance of finite-time empirical averages might not always be trivially controlled by the asymptotic variance. For one thing, the lower bound in (24) is vacuous when the asymptotic variance vanishes. We saw that the empirical standard deviation can indeed decay faster at the order of t^{-1} in Example 4. Moreover, the upper bound in (25) can fail without a positive singular-value gap: see Example 5 below. This is in contrast to the reversible scenario, where the finite-time variance is bounded by the asymptotic variance whenever the latter is finite. Since this bound has no direct reference to the spectral gap, we can get a uniform upper bound for

finite-time variance of a class of observables if we can bound their asymptotic variance uniformly. The aforementioned bound can be quickly deduced from (16) and the observation that $P_{r/2} = P_{r/2}^*$ for $r > 0$ when the process is reversible (cf. Proposition 4.29 in [AF14] and Lemma 2.3 in [CCG12]):

$$\begin{aligned} \text{Var}_\mu(\mu_t g) &= \frac{2}{t^2} \int_0^t (t-r) \text{Re}\langle g, P_r g \rangle dr \\ &= \frac{4}{t^2} \int_0^{t/2} (t-2s) \|P_s g\|^2 ds \end{aligned} \quad (26)$$

$$\begin{aligned} &\leq \frac{4}{t} \int_0^\infty \|P_s g\|^2 ds \\ &= \frac{\sigma_g^2}{t}, \end{aligned} \quad (27)$$

where the last equality in (27) follows from the expression of asymptotic variance obtained by multiplying (26) by t and taking the limit $t \rightarrow \infty$.

Example 5 (Pure rotation on 2d torus). *Consider pure rotation again, but this time on the 2d torus $\mathbb{T}^2$ with drift $\alpha = (1, \beta)$ for irrational β . The Fourier modes $e_k(\theta) = e^{ik \cdot \theta}$ indexed by $k \in \mathbb{Z}^2$ have eigenvalues $ik \cdot \alpha$. Write $\xi_k := k \cdot \alpha$, the frequency of the k -th mode. Since $|\xi_k|$ can be arbitrarily small, the process does not have a gap on $L_0^2(\mathbb{T}^2)$.*

Since the dynamics is deterministic and the normalized Lebesgue measure μ is invariant, we have $\text{Var}_\mu(\mu_t g) = \|A_t g\|^2$ as in (21). Using $P_s e_k(\theta) = e^{i\xi_k s} e_k$,

$$A_t e_k = \frac{1}{t} \int_0^t P_s e_k ds = \frac{e^{i\xi_k t} - 1}{i\xi_k t} e_k.$$

We therefore get

$$\text{Var}_\mu(\mu_t e_k) = \left| \frac{e^{i\xi_k t} - 1}{i\xi_k t} \right|^2 = \frac{4 \sin^2(\xi_k t/2)}{\xi_k^2 t^2} \implies \sigma_{e_k}^2 = \lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t e_k) = 0$$

However, for each $t > 0$, we can choose k such that $t|\xi_k| \ll 1$, so that

$$\text{Var}_\mu(\mu_t e_k) = \left\| \frac{1}{t} \int_0^t e^{i\xi_k s} e_k ds \right\|^2 \approx 1.$$

Therefore, $\Delta_t = 1$ for every t . This example shows that asymptotic variance alone cannot control finite-time empirical variance uniformly over an infinite-dimensional observable class. Here, every fixed Fourier mode has zero asymptotic variance, but the frequencies accumulate at zero, so for each fixed t we can choose a mode that has barely moved. In particular, the infinite-time limit does not commute with the worst-case supremum over observables:

$$\sup_{k \neq 0} \lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t e_k) = 0, \quad \text{while} \quad \lim_{t \rightarrow \infty} \sup_{k \neq 0} \text{Var}_\mu(\mu_t e_k) = 1.$$

3. ORNSTEIN–UHLENBECK PROCESS WITH ROTATION

In the following example, the generator is normal despite not being self-adjoint in $L^2(\mu)$. Moreover, the spectrum and eigenfunctions are explicit, so we get exact global and local singular-value gaps on eigensubspaces. In this case, the variance of finite-time empirical averages is computable in the eigenbasis, so our theory in section 2 above strictly speaking does not provide more information. Nevertheless, the theory helps us obtain clean bounds for natural classes of observables, and this example serves as a good illustration for various aspects of the bounds we derived. Let

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and consider the diffusion on $\mathbb{R}^2$ with a rotation of strength α :

$$dX_t = (-X_t + \alpha J X_t) dt + \sqrt{2} dW_t, \quad \alpha \in \mathbb{R}.$$

The invariant measure is the standard Gaussian measure

$$\mu(dx) = \frac{1}{2\pi} e^{-|x|^2/2} dx = \frac{1}{2\pi} e^{-r^2/2} r dr d\theta.$$

The generator can be decomposed into symmetric and antisymmetric parts:

$$L_\alpha = \Delta - x \cdot \nabla + \alpha \partial_\theta = S + \alpha A,$$

where the symmetric part is the usual reversible Ornstein–Uhlenbeck generator $S = \Delta - x \cdot \nabla$, and the antisymmetric part is $\alpha A = \alpha \partial_\theta$. Since S is rotationally invariant, S commutes with A . L_α is then normal, and in fact its eigenfunctions can be written explicitly in terms of generalized Laguerre polynomials. See Appendix A.2 for details. We have $(\psi_{k,m})_{k \geq 0, m \in \mathbb{Z}}$ an orthonormal basis of $L^2(\mu)$, such that

$$S\psi_{k,m} = -(2k + |m|)\psi_{k,m}, \quad \partial_\theta \psi_{k,m} = im\psi_{k,m}, \quad (28)$$

so that

$$L_\alpha \psi_{k,m} = (-(2k + |m|) + i\alpha m)\psi_{k,m}.$$

This decomposition is a polar-coordinate version of the Hermite chaos decomposition for the Gaussian measure. See [CL14] for a related treatment of complex OU operators.

3.1. Explicit variance computation. Since the eigenbasis is explicit, one can also compute the empirical variance mode by mode. Let

$$n = 2k + |m|, \quad \omega = \alpha m.$$

Then

$$L_\alpha \psi_{k,m} = (-n + i\omega)\psi_{k,m}.$$

For a normalized eigenfunction $\psi_{k,m}$, the identity (15) gives

$$\text{Var}_\mu(\mu_t \psi_{k,m}) = \frac{2}{t^2} \text{Re} \left[\frac{e^{(-n+i\omega)t} - 1 - (-n+i\omega)t}{(-n+i\omega)^2} \right]. \quad (29)$$

In particular,

$$\sigma_{\psi_{k,m}}^2 = \lim_{t \rightarrow \infty} t \operatorname{Var}_{\mu}(\mu_t \psi_{k,m}) = \frac{2n}{n^2 + \omega^2} = \frac{2(2k + |m|)}{(2k + |m|)^2 + \alpha^2 m^2}. \quad (30)$$

For any observable $g \in L_0^2(\mu)$, if we can write it in the eigenbasis as $g = \sum_{k,m} c_{k,m} \psi_{k,m}$, we get the empirical variance explicitly

$$\operatorname{Var}_{\mu}(\mu_t g) = \sum_{k,m} |c_{k,m}|^2 \operatorname{Var}_{\mu}(\mu_t \psi_{k,m})$$

3.2. Uniform acceleration for nonradial observables. Since L_{α} is normal, the singular values of L_{α} are the moduli of its eigenvalues. On $L_0^2(\mu)$, the constant mode $(k, m) = (0, 0)$ is excluded, so the global singular-value gap is

$$\gamma = \inf_{(k,m) \neq (0,0)} |-(2k + |m|) + i\alpha m| = \min\{\sqrt{1 + \alpha^2}, 2\},$$

where the modes $(k, m) = (0, \pm 1)$ give $\sqrt{1 + \alpha^2}$, while the radial mode $(k, m) = (1, 0)$ gives 2. Consequently, the global singular-value gap is capped at 2 when the antisymmetric rotation strength α grows. The obstruction is the radial quadratic observable

$$\psi_{1,0} \propto |x|^2 - 2,$$

whose empirical average does not converge faster when a rotation is present.

This obstruction shows that the subspace formulation of Theorem 1 could be useful when the irreversible part of the dynamics accelerates the convergence of empirical averages in certain directions but leaves some other observables invariant, in which case we may wish to separately consider different observable subspaces. In this case, to avoid the obstruction presented by $\psi_{1,0}$, consider

$$K_{\text{ang}} := \overline{\operatorname{span}\{\psi_{k,m} : m \neq 0\}} = \left\{ f \in L_0^2(\mu) : \int_0^{2\pi} f(r, \theta) d\theta = 0 \text{ for a.e. } r \right\},$$

where the second characterization above can be obtained by considering the Fourier decomposition $f(r, \theta) = \sum_{m \in \mathbb{Z}} f_m(r) e^{im\theta}$ and noting that $\psi_{k,m}$ for $k \geq 0$ form a basis for the m -th mode. Equivalently, K_{ang} is the orthogonal complement of radial functions.

$$\gamma_{\text{ang}} = \inf_{k \geq 0, m \neq 0} \sqrt{(2k + |m|)^2 + \alpha^2 m^2} = \sqrt{1 + \alpha^2}.$$

On this closed invariant subspace, Theorem 1 and Lemma 2 give

$$\operatorname{Var}_{\mu}(\mu_t g) \leq \frac{4}{t\sqrt{1 + \alpha^2}} \|g\|^2 \quad (31)$$

for every $g \in K_{\text{ang}}$. Thus, rotation gives a uniform empirical-average acceleration for all nonradial observables, clarifying, at the level of a whole observable class, the intuition that [DLP16] verifies explicitly for linear and quadratic observables.

3.3. Gaussian chaos and sharpness of variance bounds. Recall that for the standard Gaussian measure on $\mathbb{R}^2$, the n -th Gaussian chaos $\mathcal{H}_n$ is the eigenspace of the reversible Ornstein–Uhlenbeck generator $S = \Delta - x \cdot \nabla$ with eigenvalue $-n$. Equivalently, $\mathcal{H}_n$ is spanned by products of one-dimensional Hermite polynomials of total degree n . From (28) we see that

$$\mathcal{H}_n = \text{span}\{\psi_{k,m} : 2k + |m| = n\}.$$

In particular, for a given n , the possible angular frequencies $m = -n, -n+2, \dots, n-2, n$ give $n+1$ modes, which equals $\dim \mathcal{H}_n$. We spell out the first few eigenspaces, which can be obtained from the explicit form of $\psi_{k,m}$ derived in Appendix A.2:

$$\begin{aligned}\mathcal{H}_1 &= \text{span}\{\psi_{0,1}, \psi_{0,-1}\} = \text{span}\{x_1, x_2\}, \\ \mathcal{H}_2 &= \text{span}\{\psi_{1,0}, \psi_{0,2}, \psi_{0,-2}\} = \text{span}\{|x|^2 - 2, x_1^2 - x_2^2, 2x_1x_2\}.\end{aligned}$$

In words, $\mathcal{H}_1$ is the space of centered linear observables and $\mathcal{H}_2$ is the space of centered quadratic observables. We compute the gap of L_α on the n -th Gaussian chaos:

$$\gamma_{\mathcal{H}_n} = \min_{m=-n, -n+2, \dots, n} \sqrt{n^2 + \alpha^2 m^2} = \begin{cases} n & n \text{ even} \\ \sqrt{n^2 + \alpha^2} & n \text{ odd.} \end{cases}$$

For even n , the obstruction is again the radial sector $m = 0$. If we restrict attention to the angular part of $\mathcal{H}_n$, the local gap increases with α as

$$\gamma_{\mathcal{H}_n \cap K_{\text{ang}}} = \begin{cases} \sqrt{n^2 + 4\alpha^2} & n \text{ even nonzero} \\ \sqrt{n^2 + \alpha^2} & n \text{ odd.} \end{cases}$$

Similarly, we may use (30) to compute the worst-case asymptotic variance defined in (22):

$$\begin{aligned}\sigma_{\mathcal{H}_n}^2 &= \max_{m=-n, -n+2, \dots, n} \frac{2n}{n^2 + \alpha^2 m^2} = \begin{cases} \frac{2}{n} & n \text{ even} \\ \frac{2n}{n^2 + \alpha^2} & n \text{ odd} \end{cases} \\ \sigma_{\mathcal{H}_n \cap K_{\text{ang}}}^2 &= \begin{cases} \frac{2n}{n^2 + 4\alpha^2} & n \text{ even} \\ \frac{2n}{n^2 + \alpha^2} & n \text{ odd} \end{cases}\end{aligned}$$

In particular, for fixed n , the asymptotic variance of observables in $\mathcal{H}_n \cap K_{\text{ang}}$ is of order α^{-2} . The upper bound in Theorem 1 gives

$$\text{Var}_\mu(\mu_t g) \leq \frac{4}{t\sqrt{n^2 + \alpha^2}} \|g\|^2$$

for all $g \in \mathcal{H}_n \cap K_{\text{ang}}$, which gives a weaker $|\alpha|^{-1}$ dependence as $|\alpha| \rightarrow \infty$. On the other hand, Theorem 3 gives the upper bound

$$(\Delta_t^{\mathcal{H}_n \cap K_{\text{ang}}})^2 \leq \frac{\sigma_{\mathcal{H}_n \cap K_{\text{ang}}}^2}{t} + \frac{4}{\gamma_{\mathcal{H}_n \cap K_{\text{ang}}}^2 t^2} \leq \frac{2n}{(n^2 + \alpha^2)t} + \frac{4}{(n^2 + \alpha^2)t^2},$$

which matches the α^{-2} factor of variance reduction as $|\alpha| \rightarrow \infty$.

On the larger nonradial subspace K_{ang} , the upper bound in Theorem 1 already sharply captures the α -dependence of variance reduction. Indeed, using (30) as $|\alpha| \rightarrow \infty$,

$$\sigma_{K_{\text{ang}}}^2 = \sup_{\substack{k \geq 0 \\ m \neq 0}} \frac{2(2k + |m|)}{(2k + |m|)^2 + \alpha^2 m^2} \asymp \frac{1}{|\alpha|} \asymp \frac{1}{\gamma_{\text{ang}}}.$$

The upper bound above follows from

$$\frac{2n}{n^2 + \alpha^2 m^2} \leq \frac{2n}{n^2 + \alpha^2} \leq \frac{1}{|\alpha|} \quad (m \neq 0),$$

and the matching lower bound is obtained by taking $|m| = 1$ and choosing $n = 2k + 1$ of order $|\alpha|$. Therefore, for the infinite-dimensional observable class K_{ang} , the α -dependence in the finite-time variance bound of Theorem 1 matches the reduction factor for asymptotic variance.

3.4. Practical implication for variance reduction. In higher dimensions, we have more freedom to choose the perturbation J while maintaining the invariance of the Gaussian measure. It has been widely observed in the context of Markov chains and diffusion processes that the addition of an irreversible perturbation accelerates convergence of the law of X_t and, respectively, convergence of empirical averages. It is therefore of interest to find an optimal perturbation to accelerate convergence.

The observable-class perspective we take here suggests an optimization problem different from the one emphasized in [DLP16]. There, the perturbation is chosen to minimize the asymptotic variance of a fixed observable, and their Section 4.2 solves this problem for linear observables under a Frobenius norm constraint on J . In the present framework, one can instead fix a closed invariant class of observables K , such as $\mathcal{H}_1$, $\mathcal{H}_n$, or a finite sum of chaoses, and choose J to maximize the local singular-value gap $\gamma_K(J)$, or, often equivalently, use Theorem 3 to minimize the worst-case asymptotic variance $\sigma_K^2(J)$. Solving this optimization problem gives a uniform convergence guarantee for a class of observables and extends naturally beyond linear and quadratic functions.

Example 6 (Uniform variance reduction for linear observables). *We illustrate the observable-class variance reduction perspective in a simple higher-dimensional Gaussian example. Let $\mu = \mathcal{N}(0, I_d)$, and let $J \in \mathbb{R}^{d \times d}$ be skew-symmetric. Consider*

$$L_{J,\alpha} = \Delta - x \cdot \nabla + \alpha Jx \cdot \nabla, \quad \alpha \in \mathbb{R}.$$

which leaves $\mathcal{N}(0, I_d)$ invariant. Let

$$\mathcal{H}_1 = \{f_\ell(x) = \ell \cdot x : \ell \in \mathbb{R}^d\}$$

be the first Gaussian chaos, i.e. the space of centered linear observables. Note that

$$L_{J,\alpha} f_\ell = -f_{(I+\alpha J)\ell}.$$

Indeed,

$$\Delta f_\ell = 0, \quad -x \cdot \nabla f_\ell = -\ell \cdot x, \quad (Jx) \cdot \nabla f_\ell = \ell \cdot Jx = (J^T \ell) \cdot x = -(J\ell) \cdot x.$$

Since $\|f_\ell\| = \|\ell\|_{\mathbb{R}^d}$, the singular-value gap of $L_{J,\alpha}$ on $\mathcal{H}_1$ is given by the smallest singular-value of $I + \alpha J$. Since $J^T = -J$, $(I + \alpha J)^T(I + \alpha J) = I + \alpha^2 J^T J$. Hence

$$\gamma_{\mathcal{H}_1}(J) = \sqrt{1 + \alpha^2 \lambda_{\min}(J^T J)}$$

The worst-case asymptotic variance on $\mathcal{H}_1$ can also be computed explicitly. Since

$$-L_{J,\alpha}^{-1} f_\ell = f_{(I+\alpha J)^{-1}\ell},$$

we get

$$\sigma_{f_\ell}^2 = 2 \operatorname{Re}\langle f_\ell, f_{(I+\alpha J)^{-1}\ell} \rangle = 2 \ell^T (I + \alpha J)^{-1} \ell = 2 \ell^T (I + \alpha^2 J^T J)^{-1} \ell,$$

where the last equality follows by considering the symmetric part of the quadratic form $(I + \alpha J)^{-1}$. Taking the supremum over $\|f_\ell\| = \|\ell\|_{\mathbb{R}^d} = 1$, we obtain

$$\sigma_{\mathcal{H}_1}^2(J) = \frac{2}{1 + \alpha^2 \lambda_{\min}(J^T J)}.$$

Thus, for the class of centered linear observables, maximizing the local singular-value gap $\gamma_{\mathcal{H}_1}(J)$ is equivalent to minimizing the worst-case asymptotic variance $\sigma_{\mathcal{H}_1}^2(J)$.

Therefore, variance reduction for centered linear observables over J amounts to maximizing $\lambda_{\min}(J^T J)$. If we impose the constraint in Frobenius norm $\|J\|_F = 1$, then for even d we get

$$\lambda_{\min}(J^T J) \leq \frac{1}{d} \operatorname{Tr}(J^T J) = \frac{1}{d} \|J\|_F^2 = \frac{1}{d}.$$

The equality holds precisely when all eigenvalues of $J^T J$ are equal, i.e. when $J^T J = \frac{1}{d} I$. One such optimizer is

$$J_* = \frac{1}{\sqrt{d}} \begin{pmatrix} 0 & -1 & & & \\ 1 & 0 & & & \\ & & 0 & -1 & \\ & & 1 & 0 & \\ & & & & \ddots \end{pmatrix},$$

with $d/2$ identical 2×2 rotation blocks. For this choice,

$$\gamma_{\mathcal{H}_1}(J_*) = \sqrt{1 + \frac{\alpha^2}{d}}, \quad \sigma_{\mathcal{H}_1}^2(J_*) = \frac{2}{1 + \alpha^2/d}.$$

By Theorem 3, for all $g \in \mathcal{H}_1$,

$$\operatorname{Var}_\mu(\mu_t g) \leq \left(\frac{2}{(1 + \alpha^2/d)t} + \frac{4}{(1 + \alpha^2/d)t^2} \right) \|g\|^2.$$

If d is odd, every real skew-symmetric J has a nontrivial kernel, so $\lambda_{\min}(J^T J) = 0$. In this case, no skew-symmetric perturbation J can improve the worst-case empirical or asymptotic variance over all centered linear observables:

$$\gamma_{\mathcal{H}_1}(J) = 1, \quad \sigma_{\mathcal{H}_1}^2(J) = 2.$$

This obstruction disappears if one restricts to the subspace K of linear observables f_ℓ with ℓ orthogonal to $\ker J$. One may check that K is invariant, so we can obtain variance reduction similar to the even-dimensional case.

4. RELATION TO OTHER MIXING QUANTITIES

4.1. Relation to the reversibilized process. When we have a bounded Markov generator L , its symmetric part can be written as

$$L_s := \frac{L + L^*}{2},$$

which is self-adjoint on $L^2(\mu)$. The Markov process with generator L_s is reversible. We denote by γ_s the spectral gap of L_s , which coincides with its singular-value gap. If L is in addition normal, γ is the gap of the spectrum of $L_0 = L|_{L_0^2(\mu)}$ and γ_s is the gap of the real part of the spectrum of L_0 , so it is natural to expect that γ_s lower bounds γ .

For unbounded L , the operator expression $(L + L^*)/2$ may have domain issues. In this case, we define γ_s through the symmetric form associated with L :

$$\gamma_s := \inf_{f \in \text{Dom}(L) \cap L_0^2(\mu), \|f\|=1} \text{Re}\langle f, -Lf \rangle. \quad (32)$$

This agrees with the previous definition when L is bounded. Moreover, if there is

$$\mathcal{C} \subset \text{Dom}(L) \cap \text{Dom}(L^*)$$

that is a core for L in the graph norm with $\mathbf{1} \in \mathcal{C}$, then γ_s can be computed on $\mathcal{C}$ as

$$\gamma_s = \inf_{f \in \mathcal{C} \cap L_0^2(\mu), \|f\|=1} \left\langle f, -\frac{L + L^*}{2} f \right\rangle.$$

In the smooth diffusion examples considered in the present paper, this is the usual way to understand the symmetrized generator and compute the symmetrized gap.

Proposition 3. *Let γ be the singular-value gap of L and γ_s be defined as in (32). Then*

$$\gamma_s \leq \gamma.$$

Proof. For any $f \in \text{Dom}(L) \cap L_0^2(\mu)$, we have

$$\text{Re}\langle f, -Lf \rangle \leq |\langle f, -Lf \rangle| \leq \|f\| \|Lf\|.$$

Taking the infimum over $\|f\| = 1$, we get $\gamma_s \leq \gamma$. $\square$

Remark 5. *In the discrete time case, let γ_A be the singular-value gap of the chain with transition matrix $A = \frac{1}{2}(P + P^*)$, as considered in [Cha25]. Then the same proof as above shows that $\gamma_A \leq \gamma$, improving the $\frac{1}{2}$ factor in Theorem 1.3 of [Cha25].*

It is natural to ask whether an upper bound of γ could be also obtained through γ_s . Such a bound cannot hold in the most general setting, as demonstrated by Example 1. That is, diffusion on the unit circle with generator $L = \partial_\theta^2 + \alpha \partial_\theta$ has $\gamma_s = 1$ but $\gamma = \sqrt{1 + \alpha^2}$. From this counterexample it is natural to expect that if we demand the strength of the dynamics (in this case the drift) to be bounded, then γ might be bounded above in terms of γ_s . We prove such a bound for jump processes with bounded jump rates.

Let S be a measurable space. Following the convention of [LS88], we consider unnormalized jump kernel $q(x, dy)$: for each x , $q(x, \cdot)$ is a finite measure on S , and for each measurable A , the map $x \mapsto q(x, A)$ is measurable. The total mass

$$q(x) := q(x, S)$$

is the jump rate at x . The interpretation is that $q(x)$ is the exponential rate at which a jump is proposed from state x , and, conditional on such a jump, the next state is sampled from the normalized kernel

$$p(x, dy) := \frac{q(x, dy)}{q(x)} \quad \text{if } q(x) > 0.$$

The generator for this process is

$$Lf(x) = \int (f(y) - f(x)) q(x, dy). \quad (33)$$

Since the jump rates are bounded and μ is invariant, L is a bounded operator on $L^p(\mu)$ for every $1 \leq p \leq \infty$. Indeed, for $r \geq q_{\max}$,

$$\Pi(x, dy) = \frac{q(x, dy)}{r} + \left(1 - \frac{q(x)}{r}\right) \delta_x(dy)$$

is a Markov kernel preserving μ , and $L = r(\Pi - I)$. Therefore, $\|\Pi\|_{L^p \rightarrow L^p} \leq 1$, hence

$$\|Lf\|_p \leq 2r\|f\|_p \quad \text{for } 1 \leq p \leq \infty.$$

In particular, $\text{Dom}_{L^2}(L) = L^2(\mu)$.

Proposition 4. *Consider a jump process on a measurable space S with generator (33) and an invariant probability measure μ . Assume the jump rates are bounded:*

$$q(x, S) \leq q_{\max} < \infty \quad \text{for } \mu\text{-a.e. } x.$$

Let γ be the singular-value gap of L on $L_0^2(\mu)$, and let γ_s be the spectral gap of the symmetrized generator L_s , which is a bounded self-adjoint operator. Then

$$\gamma \leq \sqrt{2q_{\max}\gamma_s}.$$

Proof. Let

$$Q(dx, dy) := \mu(dx)q(x, dy)$$

be the stationary jump flux. Since μ is invariant, for every bounded measurable h ,

$$\int Lh(x) \mu(dx) = 0.$$

By truncation, the identity also holds for every $h \in L^1(\mu)$, so

$$\iint h(y) Q(dx, dy) = \iint h(x) Q(dx, dy). \quad (34)$$

For $f \in L^2(\mu)$, the Dirichlet form of the symmetrized process is $\langle f, -L_s f \rangle = \text{Re} \langle f, -L f \rangle$. We claim that

$$\langle f, -L_s f \rangle = \frac{1}{2} \iint |f(y) - f(x)|^2 Q(dx, dy). \quad (35)$$

Indeed,

$$\begin{aligned} 2 \text{Re} \langle f, -L f \rangle &= 2 \text{Re} \int \overline{f(x)} \int (f(x) - f(y)) q(x, dy) \mu(dx) \\ &= 2 \iint |f(x)|^2 Q(dx, dy) - 2 \text{Re} \iint \overline{f(x)} f(y) Q(dx, dy). \end{aligned}$$

Using (34) with $h = |f|^2$,

$$\iint |f(x)|^2 Q(dx, dy) = \iint |f(y)|^2 Q(dx, dy).$$

Therefore

$$\begin{aligned} 2 \text{Re} \langle f, -L f \rangle &= \iint \left(|f(x)|^2 + |f(y)|^2 - 2 \text{Re}(\overline{f(x)} f(y)) \right) Q(dx, dy) \\ &= \iint |f(y) - f(x)|^2 Q(dx, dy). \end{aligned}$$

This proves the claimed identity (35). Next, by Cauchy–Schwarz,

$$\begin{aligned} |L f(x)|^2 &= \left| \int (f(y) - f(x)) q(x, dy) \right|^2 \\ &\leq q(x, S) \int |f(y) - f(x)|^2 q(x, dy) \\ &\leq q_{\max} \int |f(y) - f(x)|^2 q(x, dy). \end{aligned}$$

Integrating over x gives

$$\|L f\|^2 \leq q_{\max} \iint |f(y) - f(x)|^2 Q(dx, dy) = 2 q_{\max} \langle f, -L_s f \rangle.$$

Taking a sequence $(f_n)_{n \geq 1}$ in $L_0^2(\mu)$ such that $\|f_n\| = 1$ and $\langle f_n, -L_s f_n \rangle \rightarrow \gamma_s$,

$$\gamma^2 = \inf_{\substack{f \in L_0^2(\mu) \\ \|f\|=1}} \|L f\|^2 \leq \liminf_{n \rightarrow \infty} \|L f_n\|^2 \leq \liminf_{n \rightarrow \infty} 2 q_{\max} \langle f_n, -L_s f_n \rangle = 2 q_{\max} \gamma_s,$$

as desired. $\square$

4.2. Relation to the mixing time. Next, we relate the singular-value gap γ to the mixing time in total-variation distance. For a Markov process with an invariant probability measure μ , write

$$d(t) := \sup_{\nu \in \mathcal{P}(S)} \|\nu P_t - \mu\|_{\text{TV}}, \quad t_{\text{mix}}(\varepsilon) := \inf\{t \geq 0 : d(t) \leq \varepsilon\},$$

where $\mathcal{P}(S)$ is the set of probability measures on S . Equivalently,

$$d(t) = \sup_{x \in S} \|P_t(x, \cdot) - \mu\|_{\text{TV}}.$$

Indeed, the inequality “ $\geq$ ” follows by taking $\nu = \delta_x$. Conversely, for any measurable A ,

$$(\nu P_t - \mu)(A) = \int (P_t(x, A) - \mu(A)) \nu(dx).$$

Therefore

$$|(\nu P_t - \mu)(A)| \leq \sup_{x \in S} |P_t(x, A) - \mu(A)| \leq \sup_{x \in S} \|P_t(x, \cdot) - \mu\|_{\text{TV}}.$$

Taking the supremum over A proves the reverse inequality.

The above definitions are canonical for finite state space discrete-time chains, where t takes values in nonnegative integers and $P_t = P^t$. On a general state space, total variation is one of many metrics for mixing, and it is usually more meaningful when the process is uniformly ergodic in the sense that $d(t) \rightarrow 0$ as $t \rightarrow \infty$.

4.2.1. Lower bound on $t_{\text{mix}}(\varepsilon)$. For a Markov kernel $\pi(x, dy)$ on a standard Borel space S , write

$$\delta(\pi) := \sup_{x, x' \in S} \|\pi(x, \cdot) - \pi(x', \cdot)\|_{\text{TV}}.$$

Note that $\delta(P_t) \leq 2d(t)$ by triangular inequality. We need the following lemma from [SV05]. The original paper considers real-valued functions, but the proof still goes through for complex-valued functions.

Lemma 3 (Complex form of Lemma 4.1 in [SV05]). *Let S be a standard Borel space, Q a probability measure on $S \times S$ with marginals α and β , and $\pi(x, dy)$ a Markov kernel satisfying $Q(dx, dy) = \alpha(dx)\pi(x, dy)$. If $f \in L^2(\alpha)$ and $g \in L^2(\beta)$ satisfy*

$$\int f d\alpha = \int g d\beta = 0,$$

then

$$\left| \iint f(x)g(y) Q(dx, dy) \right| \leq \sqrt{\delta(\pi)} \|f\|_{L^2(\alpha)} \|g\|_{L^2(\beta)}$$

Corollary 3. *Let $\pi(x, dy)$ be a Markov kernel with an invariant probability measure μ . As usual, use $\|\cdot\|$ to denote the $L^2(\mu)$ -norm. Then for any function $g \in L_0^2(\mu)$,*

$$\|\pi g\| \leq \sqrt{\delta(\pi)} \|g\|.$$

Proof. Since μ is invariant, $\mu(\pi g) = \mu(g) = 0$. We can then apply Lemma 3 with $\alpha = \beta = \mu$ and $f = \overline{\pi g}$ to get

$$\begin{aligned} \|\pi g\|^2 &= \int \overline{\pi g}(x) \pi g(x) \mu(dx) \\ &= \iint \overline{\pi g}(x) g(y) \pi(x, dy) \mu(dx) \\ &\leq \sqrt{\delta(\pi)} \|\pi g\| \|g\|. \end{aligned}$$

If $\pi g = 0$, we are done; otherwise, divide both sides by $\|\pi g\|$. □

Theorem 4. Consider either a continuous-time Markov process or a discrete-time Markov chain defined on a standard Borel space, with singular-value gap γ defined in (5) and (4) respectively. Let $\varepsilon \in (0, \frac{1}{2})$. Let $t > 0$ be such that $d(t) \leq \varepsilon$. Then $\gamma \geq \frac{1-\sqrt{2\varepsilon}}{t}$. In particular,

$$\frac{1-\sqrt{2\varepsilon}}{\gamma} \leq t_{\text{mix}}(\varepsilon).$$

Proof. Fix $\varepsilon \in (0, \frac{1}{2})$ and $f \in \text{Dom}(L) \cap L_0^2(\mu)$. The idea is that to obtain a lower bound on γ , we can lower bound $\|Lf\|$ by $\|f\|$ through the intermediate quantity $\|f - P_t f\|$ and control $\|P_t f\|$ by contraction in total-variation distance, thanks to Corollary 3.

Continuous-time case: Let $t > 0$ be such that $d(t) \leq \varepsilon$. Using the identity $f - P_t f = \int_0^t P_s(-Lf) ds$, we have

$$\|f - P_t f\| \leq \int_0^t \|P_s Lf\| ds \leq t \|Lf\|. \quad (36)$$

Recall that $\delta(P_t) \leq 2d(t) \leq 2\varepsilon$. Apply Corollary 3 with $\pi = P_t$ to get

$$\|P_t f\| \leq \sqrt{\delta(P_t)} \|f\| \leq \sqrt{2\varepsilon} \|f\|$$

Therefore,

$$\|f - P_t f\| \geq \|f\| - \|P_t f\| \geq (1 - \sqrt{2\varepsilon}) \|f\|. \quad (37)$$

Combining (36) and (37), we get

$$\frac{\|Lf\|}{\|f\|} \geq \frac{1 - \sqrt{2\varepsilon}}{t}.$$

Taking infimum over $f \in L_0^2(\mu)$, we conclude $\gamma \geq \frac{1-\sqrt{2\varepsilon}}{t}$. We get $t_{\text{mix}}(\varepsilon) \geq \frac{1-\sqrt{2\varepsilon}}{\gamma}$ by taking infimum over all t such that $d(t) \leq \varepsilon$.

Discrete-time case: Let $n \geq 1$ be such that $d(n) \leq \varepsilon$. Applying Corollary 3 with $\pi = P^n$, the same argument leading to (37) gives

$$\|f - P^n f\| \geq (1 - \sqrt{2\varepsilon}) \|f\|. \quad (38)$$

On the other hand, with $L = P - I$,

$$\|f - P^n f\| \leq \sum_{k=0}^{n-1} \|P^{k+1} f - P^k f\| = \sum_{k=0}^{n-1} \|P^k (P - I) f\| \leq n \|Lf\| \quad (39)$$

Combining (38) and (39), we get

$$\frac{\|Lf\|}{\|f\|} \geq \frac{1 - \sqrt{2\varepsilon}}{n}.$$

Taking infimum over $f \in L_0^2(\mu)$, we conclude $\gamma \geq \frac{1-\sqrt{2\varepsilon}}{n}$. In particular, we can take $n = t_{\text{mix}}(\varepsilon)$. $\square$

Remark 6. The discrete-time case above improves by a constant the lower bound on the mixing time in Theorem 1.4 of [Cha25]. In fact, a bound with a sharper constant than stated in Theorem 1.4 of [Cha25] could be derived from an observation in the same paper: Let γ_{ps} be the pseudo-spectral gap in [Pau18], defined by

$$\gamma_{\text{ps}} := \max_{k \geq 1} \frac{\gamma((P^*)^k P^k)}{k},$$

where $\gamma((P^*)^k P^k)$ is the spectral gap of the self-adjoint $(P^*)^k P^k$. Proposition 3.4 in [Pau18] shows that for $\varepsilon \in (0, \frac{1}{2})$, $\frac{1-2\varepsilon}{\gamma_{\text{ps}}} \leq t_{\text{mix}}(\varepsilon)$. As observed in [Cha25], $\gamma \geq \frac{1}{2}\gamma_{\text{ps}}$. These two estimates combine to give

$$\frac{1-2\varepsilon}{2\gamma} \leq t_{\text{mix}}(\varepsilon)$$

in the discrete-time case, with a worse constant compared to what we have above.

4.2.2. Upper bound on $t_{\text{mix}}(\varepsilon)$. We can obtain an upper bound on the mixing time by γ only in the restrictive setting of jump processes on finite state space. Note from the proof below that the upper bound can in fact be stated in terms of γ_s .

Proposition 5. Consider a jump process on a finite state space S with an invariant probability measure μ ; the generator is given as in (33). Assume $\gamma > 0$,

$$\mu_{\min} := \min_{x \in S} \mu(x) > 0,$$

and that the jump rates are bounded by $q_{\max}$. Then, for $\varepsilon \in (0, 1/2)$,

$$t_{\text{mix}}(\varepsilon) \leq \frac{2q_{\max}}{\gamma^2} \log \left(\frac{1}{2\varepsilon\sqrt{\mu_{\min}}} \right).$$

Proof. Let γ_s be the symmetrized gap on $L_0^2(\mu)$ defined in (32). We first note that γ_s controls the $L^2(\mu)$ -decay of the original semigroup, with a derivation standard in the reversible scenario. If $f \in L_0^2(\mu)$, then $P_t f \in L_0^2(\mu)$ because μ is invariant. Therefore

$$\begin{aligned} \frac{d}{dt} \|P_t f\|^2 &= \langle LP_t f, P_t f \rangle + \langle P_t f, LP_t f \rangle \\ &= 2 \operatorname{Re} \langle P_t f, LP_t f \rangle \\ &\leq -2\gamma_s \|P_t f\|^2. \end{aligned}$$

By Gronwall's inequality,

$$\|P_t f\| \leq e^{-\gamma_s t} \|f\|. \quad (40)$$

Next, recall the dual characterization of total variation: for any $x \in S$,

$$\|P_t(x, \cdot) - \mu\|_{\text{TV}} = \frac{1}{2} \sup_{\|h\|_{\infty} \leq 1} |P_t h(x) - \mu h|.$$

For any bounded measurable h , let $\bar{h} := h - \mu h$. Then $\bar{h} \in L_0^2(\mu)$, and

$$P_t h(x) - \mu h = P_t \bar{h}(x).$$

Moreover, if $\|h\|_\infty \leq 1$, then $\|\bar{h}\| \leq 1$. For any $g \in L^2(\mu)$,

$$\sqrt{\mu(x)} |g(x)| \leq \left(\sum_{y \in S} \mu(y) |g(y)|^2 \right)^{1/2} = \|g\|$$

Combining the above ingredients, we get for any h on S with $\|h\|_\infty \leq 1$,

$$|P_t h(x) - \mu h| = |P_t \bar{h}(x)| \leq \frac{1}{\sqrt{\mu(x)}} e^{-\gamma_s t} \|\bar{h}\| \leq \frac{1}{\sqrt{\mu_{\min}}} e^{-\gamma_s t}.$$

Recall from Proposition 4 that

$$\gamma_s \geq \frac{\gamma^2}{2q_{\max}}.$$

Hence

$$\|P_t(x, \cdot) - \mu\|_{\text{TV}} \leq \frac{1}{2\sqrt{\mu_{\min}}} e^{-\gamma_s t} \leq \frac{1}{2\sqrt{\mu_{\min}}} \exp\left(\frac{-\gamma^2 t}{2q_{\max}}\right).$$

In particular, for $t = \frac{2q_{\max}}{\gamma^2} \log\left(\frac{1}{2\varepsilon\sqrt{\mu_{\min}}}\right)$, we have $\|P_t(x, \cdot) - \mu\|_{\text{TV}} \leq \varepsilon$. $\square$

The finite state space assumption in Proposition 5 above is essential. On a countably infinite state space, a positive spectral gap does not imply finite worst-case total-variation mixing time, even for a reversible jump process with bounded rates.

Example 7. Consider the continuous-time $M/M/1$ queue, or reflected biased random walk, on $\mathbb{N} = \{0, 1, 2, \dots\}$. Fix $0 < \lambda < \mu$, and let

$$q(n, n+1) = \lambda, \quad n \geq 0 \quad \text{and} \quad q(n, n-1) = \mu, \quad n \geq 1.$$

The generator of this process is given by

$$\begin{aligned} (Lf)(0) &= \lambda(f(1) - f(0)), \\ (Lf)(n) &= \lambda(f(n+1) - f(n)) + \mu(f(n-1) - f(n)), \quad n \geq 1. \end{aligned}$$

The total jump rate is bounded by $\lambda + \mu$. The invariant measure is the geometric distribution

$$\pi(n) = (1 - \rho)\rho^n, \quad \rho = \frac{\lambda}{\mu} < 1.$$

Moreover, $\pi(n)\lambda = \pi(n+1)\mu$, so the process is reversible. This process has a positive $L^2(\pi)$ spectral gap. In fact, we can explicitly compute the spectral gap to be (see Appendix A.3)

$$\gamma = (\sqrt{\mu} - \sqrt{\lambda})^2.$$

Nevertheless, the worst-case total-variation distance satisfies

$$d(t) := \sup_{n \in \mathbb{N}} \|P_t(n, \cdot) - \pi\|_{\text{TV}} = 1$$

for every finite t . To see this, fix $\eta > 0$, and choose R so large that

$$\pi(\{0, 1, \dots, R\}) > 1 - \eta.$$

If the chain starts from $n > R$, then it must make at least $n - R$ downward jumps before it can enter $\{0, 1, \dots, R\}$. The total number of jumps by time t is stochastically dominated by a Poisson random variable with mean $(\lambda + \mu)t$. Hence

$$P_t(n, \{0, 1, \dots, R\}) \leq \mathbb{P}(\text{Pois}((\lambda + \mu)t) \geq n - R) \longrightarrow 0$$

as $n \rightarrow \infty$. Therefore

$$\begin{aligned} \sup_{n \in \mathbb{N}} \|P_t(n, \cdot) - \pi\|_{\text{TV}} &\geq \limsup_{n \rightarrow \infty} |P_t(n, \{0, 1, \dots, R\}) - \pi(\{0, 1, \dots, R\})| \\ &\geq 1 - \eta. \end{aligned}$$

Taking $\eta \downarrow 0$, we get $d(t) = 1$ for every finite t . Thus,

$$t_{\text{mix}}(\varepsilon) = \infty$$

for every $\varepsilon < 1$, despite the positive spectral gap.

The discrete-time version of this jump process was discussed in e.g. [KM03] as an example Markov chain that is geometrically ergodic but not Doeblin recurrent. For our purposes, this process shows that an $L^2(\pi)$ -spectral gap cannot control worst-case total-variation mixing when the process has finite “speed” while the invariant measure has unbounded support.

4.3. Relation to the Cheeger constant. Consider a stationary Markov jump process on a standard Borel space S with jump kernel $q(x, dy)$ and invariant probability measure μ . Let the jump rate be bounded by $q_{\max} := \text{ess sup}_x q(x, S) < \infty$. Recall that the generator is

$$Lf(x) = \int (f(y) - f(x)) q(x, dy). \quad (41)$$

Let

$$Q(dx, dy) := \mu(dx)q(x, dy)$$

be the stationary flux of the jump process. The Cheeger constant is defined by

$$\xi := \inf_{A \subset S: 0 < \mu(A) \leq \frac{1}{2}} \frac{Q(A, A^c)}{\mu(A)} \quad (42)$$

The constant ξ is also known as the bottleneck ratio, with several slightly different definitions in the literature. When the jump process is reversible i.e. L is self-adjoint in $L^2(\mu)$, we recall the classical result in [LS88], Theorems 2.1 and 3.5:

$$\frac{\xi^2}{2q_{\max}} \leq \gamma \leq 2\xi, \quad (43)$$

where γ is the usual spectral gap. In the general nonreversible setting, we obtain

Theorem 5. Consider a Markov jump process on a standard Borel space with generator (41), an invariant probability measure μ , and jump rates bounded by $q(x, S) \leq q_{\max} < \infty$ for μ -a.e. x . Let γ be the associated singular-value gap and ξ the Cheeger constant defined in (42). Then

$$\frac{\xi^2}{2q_{\max}} \leq \gamma \leq \frac{36}{5}\xi.$$

Before proving Theorem 5, we first record some basic characterizations of the adjoint jump process. Define the reversed flux measure by

$$Q^\dagger(dx, dy) := Q(dy, dx).$$

Since μ is invariant, for every measurable set A ,

$$0 = \int L \mathbf{1}_A d\mu = Q(S, A) - Q(A, S). \quad (44)$$

Thus, we have the equality of marginals

$$q(x) \mu(dx) = Q(dx, S) = Q(S, dx) = Q^\dagger(dx, S).$$

In particular, since the first marginal of $Q^\dagger$ is absolutely continuous with respect to μ and S is standard Borel, there exists a finite kernel $q^\dagger(x, dy)$ such that

$$Q^\dagger(dx, dy) = \mu(dx) q^\dagger(x, dy).$$

Unsurprisingly, time reversal corresponds to $L^2(\mu)$ -adjoint:

Lemma 4. Let L^* be the adjoint of L in $L^2(\mu)$. Then $q^\dagger(x, dy)$ is the jump kernel of the adjoint process with generator L^* . Consequently, the reversed flux measure $Q^\dagger(dx, dy)$ is the stationary flux of the adjoint process.

Proof. The process with jump kernel $q^\dagger(x, dy)$ has generator

$$L^\dagger g(x) := \int (g(y) - g(x)) q^\dagger(x, dy).$$

We claim $L^\dagger = L^*$. Indeed, for any $f, g \in L^2(\mu)$,

$$\begin{aligned} \langle g, Lf \rangle &= \iint \overline{g(x)} (f(y) - f(x)) Q(dx, dy) \\ &= \iint \overline{g(y)} f(x) Q^\dagger(dx, dy) - \int \overline{g(x)} f(x) Q(dx, S) \\ &= \iint \overline{g(y)} f(x) Q^\dagger(dx, dy) - \int \overline{g(x)} f(x) Q^\dagger(dx, S) \\ &= \iint (\overline{g(y)} - \overline{g(x)}) f(x) Q^\dagger(dx, dy) \\ &= \int \overline{\int (g(y) - g(x)) q^\dagger(x, dy) f(x) \mu(dx)} \\ &= \langle L^\dagger g, f \rangle. \end{aligned}$$

This shows that $q^\dagger(x, dy)$ and $Q^\dagger(dx, dy)$ are, respectively, the jump kernel and the stationary flux of the adjoint process. $\square$

We are now ready to prove Theorem 5. Note that the content of the lower bound is essentially in the proof of [LS88]. The proof of the upper bound refines the proof idea in Theorem 1.6 of [Cha25] and crucially depends on the connection between the singular-value gap and the empirical variance of observables.

Proof of the lower bound in Theorem 5. Consider the reversibilized jump process with generator $L_s = \frac{L+L^*}{2}$. Thanks to Lemma 4, we see that L_s has stationary flux

$$Q_s(dx, dy) = \frac{1}{2}(Q(dx, dy) + Q^\dagger(dx, dy)) = \frac{1}{2}(Q(dx, dy) + Q(dy, dx)).$$

Moreover, recall from (44) that for any measurable set A , $Q(S, A) = Q(A, S)$. Since $Q(S, A) = Q(A, A) + Q(A^c, A)$ and $Q(A, S) = Q(A, A) + Q(A, A^c)$, we deduce that

$$Q(A, A^c) = Q(A^c, A).$$

Consequently,

$$Q_s(A, A^c) = \frac{1}{2}(Q(A, A^c) + Q(A^c, A)) = Q(A, A^c).$$

Therefore, by the definition (42), the Cheeger constant ξ_s of the reversibilized jump process is equal to the Cheeger constant of the original process:

$$\xi_s = \xi.$$

Let γ_s be the spectral gap of the reversibilized process, satisfying $\gamma \geq \gamma_s$ by Proposition 3. Since $Q_s(dx, S) = Q(dx, S)$, the jump rates of the reversibilized process are also bounded by $q_{\max}$. By the Cheeger lower bound for reversible jump processes in (43), we conclude

$$\gamma \geq \gamma_s \geq \frac{\xi_s^2}{2q_{\max}} = \frac{\xi^2}{2q_{\max}}.$$

□

Proof of the upper bound in Theorem 5. It suffices to prove the estimate for an arbitrary measurable set $A \subset S$ with $0 < \mu(A) \leq 1/2$, and then take the infimum over such A . Write

$$a := \mu(A), \quad c := 1 - a, \quad \zeta(A) := \frac{Q(A, A^c)}{\mu(A)}.$$

If $\gamma = 0$, there is nothing to prove; if $\zeta(A) = 0$, then the argument below in particular gives $\gamma = 0$. Thus, assume $\gamma, \zeta(A) > 0$ and consider

$$g_A := \frac{\mathbf{1}_A - a}{\sqrt{a(1-a)}}.$$

Then $g_A \in L_0^2(\mu)$ and $\|g_A\| = 1$. We first obtain a lower bound on the stationary covariance of g_A . Let

$$p_s := \mathbb{P}_\mu(X_0 \in A, X_s \in A^c).$$

Since $X_0 \sim \mu$,

$$\mathbb{P}_\mu(X_0 \in A, X_s \in A) = a - p_s.$$

Therefore,

$$\mathbb{E}_\mu[g_A(X_0)g_A(X_s)] = \frac{\mathbb{P}_\mu(X_0 \in A, X_s \in A) - a^2}{a(1-a)} = 1 - \frac{p_s}{a(1-a)}.$$

The event $\{X_0 \in A, X_s \in A^c\}$ can occur only if the process makes at least one jump from A to A^c during the time interval $[0, s]$. Let $N_s(A, A^c)$ denote the number of such jumps. By Markov's inequality,

$$p_s \leq \mathbb{E}_\mu N_s(A, A^c).$$

Since the process is stationary, the expected number of jumps from A to A^c in time s is

$$\mathbb{E}_\mu N_s(A, A^c) = s Q(A, A^c).$$

Hence

$$\mathbb{E}_\mu[g_A(X_0)g_A(X_s)] \geq 1 - \frac{s Q(A, A^c)}{a(1-a)} = 1 - \frac{\zeta(A)s}{c}.$$

Using the stationary covariance identity (16), we get

$$\begin{aligned} \text{Var}_\mu(\mu_t g_A) &= \frac{2}{t^2} \int_0^t (t-s) \mathbb{E}_\mu[g_A(X_0)g_A(X_s)] ds \\ &\geq \frac{2}{t^2} \int_0^t (t-s) \left(1 - \frac{\zeta(A)s}{c}\right) ds \\ &= 1 - \frac{\zeta(A)t}{3c}. \end{aligned}$$

On the other hand, by the upper bounds in (23) (24),

$$\text{Var}_\mu(\mu_t g_A) \leq \frac{2}{\gamma t} + \frac{4}{\gamma^2 t^2}.$$

We now choose t optimally. Write

$$x := \frac{\gamma}{\zeta(A)}, \quad t = \frac{rc}{\zeta(A)}, \quad r \in (0, 3).$$

Then the lower and upper bounds imply

$$1 - \frac{r}{3} \leq \frac{2}{xcr} + \frac{4}{x^2 c^2 r^2}.$$

Let $y := cx$. Since c is fixed for a fixed set A , minimizing the constant in the upper bound amounts to minimizing y . For fixed r , the inequality above gives

$$y \leq F(r) := \frac{1 + \sqrt{1 + 4(1 - r/3)}}{r(1 - r/3)}.$$

A direct minimization shows that F is minimized at $r = 5/3$, where

$$F(5/3) = \frac{18}{5}.$$

Indeed, setting $u = 1 - r/3$ and $s = \sqrt{1 + 4u}$, one obtains

$$F(r) = \frac{16}{3(s-1)(5-s^2)},$$

so minimizing F is equivalent to maximizing $(s-1)(5-s^2)$, whose unique critical point and local maximum in $(1, \sqrt{5})$ is at $s = 5/3$, which corresponds to $r = 5/3$. Therefore, with

$$t = \frac{5c}{3\zeta(A)},$$

we get

$$\frac{\gamma}{\zeta(A)} \leq \frac{18}{5c} \leq \frac{36}{5},$$

since $c \geq 1/2$. Hence $\gamma \leq \frac{36}{5}\zeta(A)$. Taking the infimum over A proves

$$\gamma \leq \frac{36}{5}\xi.$$

□

Remark 7. The constant $1/2$ in the lower bound above improves the one stated in [Cha25] for finite-state discrete-time chains and matches the constant in the reversible case. We already improved the constant in the upper bound from 32 to $36/5$ by refining Chatterjee's proof strategy, but perhaps it could be further optimized.

5. APPLICATIONS

5.1. Shift-register chain on binary de Bruijn graph. Let

$$B_n = \{-1, 1\}^n$$

Define a discrete-time Markov chain on B_n such that from

$$x = (x_1, \dots, x_n)$$

the chain moves to

$$(x_2, \dots, x_n, +1) \quad \text{or} \quad (x_2, \dots, x_n, -1)$$

with probability $1/2$ each. This is the simple random walk on the directed binary de Bruijn graph. The chain is nonreversible and has uniform stationary measure μ . Indeed, after n steps, all original coordinates have been forgotten, so

$$P^n(x, \cdot) = \mu \quad \text{for every } x \in B_n.$$

On the other hand, for $k < n$, the law $P^k(x, \cdot)$ is uniform on only 2^k states, while μ is uniform on 2^n states. Hence

$$\|P^k(x, \cdot) - \mu\|_{\text{TV}} = 1 - 2^{k-n}, \quad 0 \leq k < n.$$

In particular, for every $\varepsilon < 1/2$,

$$t_{\text{mix}}(\varepsilon) = n.$$

We now compute the singular-value gap of $L = P - I$. Let

$$\chi_A(x) := \prod_{i \in A} x_i, \quad A \subseteq \{1, \dots, n\},$$

be the Walsh characters. These form an orthonormal basis of $L^2(\mu)$, where the constant subspace is spanned by $\chi_\emptyset = 1$; see section 1.3 of [ODo14]. For $A \neq \emptyset$,

$$P\chi_A = \begin{cases} \chi_{A+1}, & n \notin A, \\ 0, & n \in A, \end{cases}$$

where

$$A + 1 := \{i + 1 : i \in A, i < n\}.$$

Thus P decomposes $L^2_0(\mu)$ into orthogonal nilpotent shift chains. The longest such chain has length n , for example:

$$\chi_{\{1\}} \mapsto \chi_{\{2\}} \mapsto \dots \mapsto \chi_{\{n\}} \mapsto 0.$$

On a shift chain of length ℓ , P is the nilpotent shift

$$S_\ell : e_1 \mapsto e_2 \mapsto \dots \mapsto e_\ell \mapsto 0. \quad (45)$$

The smallest singular value of $I - P$ on this ℓ -dimensional subspace is the square root of the smallest eigenvalue of the tridiagonal matrix $(I - S_\ell)^*(I - S_\ell)$:

$$2 \sin \left(\frac{\pi}{4\ell + 2} \right) \asymp \frac{1}{\ell}.$$

Therefore, the global singular-value gap is

$$\gamma_n = 2 \sin \left(\frac{\pi}{4n + 2} \right) \asymp \frac{1}{n}.$$

Using Theorem 4, we get

$$t_{\text{mix}}(\varepsilon) \geq \frac{1 - \sqrt{2\varepsilon}}{\gamma_n} \asymp n,$$

which has the correct order.

This example is prototypically nonnormal and illustrates a slow-mixing mechanism caused by deterministic or nearly deterministic memory that is not captured by traditional spectral quantities. On $L^2_0(\mu)$, the transition operator is nilpotent, so all eigenvalues of P

are 0, hence all eigenvalues of $L = P - I$ are -1 . Thus, the distance between zero and the spectrum of L is 1. One can also check that the symmetrized gap γ_s has order n^{-2} . Indeed, $-L_s$ acts on a shift block of length ℓ as

$$I - \frac{S_\ell + S_\ell^*}{2},$$

with smallest eigenvalue

$$1 - \cos \frac{\pi}{\ell + 1} \asymp \frac{1}{\ell^2}.$$

By contrast, the singular-value gap of L is of order n^{-1} , exactly reflecting the length of the longest nilpotent shift chain and giving the correct order for the mixing-time lower bound. In this case, the reason that the singular-value gap “sees” the length of the shift chains can be understood by the expansion

$$(I - S_\ell)^{-1} = I + S_\ell + \cdots + S_\ell^{\ell-1},$$

from which we see that $\|(I - S_\ell)^{-1}\|$, which is the inverse of the local singular-value gap by Lemma 2, has order ℓ .

5.2. A hypoelliptic shear on the torus. We first observe a mechanism for a Markov generator to possess a positive singular-value gap.

Proposition 6 (Compact-resolvent criterion). *Let L be a Markov generator on $L^2(\mu)$, where μ is an invariant probability measure on a measurable space. Suppose that L has compact resolvent. Equivalently, the embedding of $(\text{Dom}(L), \|\cdot\| + \|L \cdot\|)$ into $L^2(\mu)$ is compact. Suppose further that the kernel of L consists only of constants. Then L has a positive singular-value gap on $L_0^2(\mu)$:*

$$\gamma := \inf_{f \in \text{Dom}(L) \cap L_0^2(\mu) \setminus \{0\}} \frac{\|Lf\|}{\|f\|} > 0.$$

Consequently, the empirical-variance bounds of Corollary 1 apply.

Proof. Since μ is invariant, $L_0^2(\mu)$ is invariant under the semigroup and hence under the resolvent. Therefore, the restriction L_0 of L to $L_0^2(\mu)$ also has compact resolvent. For an operator with compact resolvent, the spectrum consists only of isolated eigenvalues of finite algebraic multiplicity, with no finite accumulation point. Since the kernel of L consists only of constants, 0 lies in the resolvent set of L_0 . Therefore, $L_0 : L_0^2(\mu) \cap \text{Dom}(L) \rightarrow L_0^2(\mu)$ is bijective with bounded inverse, hence $\gamma > 0$. $\square$

Hypoelliptic diffusion generators often satisfy the assumptions of the previous proposition. We illustrate this point with the following example. Let $M = \mathbb{T}^2$ and consider

$$L = \partial_y^2 + \sin(y)\partial_x.$$

An invariant measure μ is the normalized Lebesgue measure. The operator is nonreversible: ∂_y^2 is symmetric and $\sin(y)\partial_x$ is skew-adjoint in $L^2(\mu)$. Unlike the OU process in section 3, the spectrum of L is not explicit.

Nevertheless, we can check that L has a positive singular-value gap on $L_0^2(\mu)$ by checking the conditions of Proposition 6. We first note that L satisfies Hörmander's bracket condition [Hör67]. The relevant vector fields are

$$X_1 = \partial_y, \quad X_0 = \sin(y)\partial_x.$$

Their commutator is

$$[X_1, X_0] = \cos(y)\partial_x.$$

Since $\sin(y)$ and $\cos(y)$ do not vanish simultaneously, the vector fields $X_1, X_0, [X_1, X_0]$ span the tangent space at every point of $\mathbb{T}^2$. Therefore, Hörmander's bracket condition holds, and L is hypoelliptic.

We now identify the kernel of L . If $Lf = 0$, then hypoellipticity implies that f is smooth. Taking real parts in the identity $\langle f, Lf \rangle = 0$, we obtain

$$0 = \operatorname{Re}\langle f, Lf \rangle = -\|\partial_y f\|^2.$$

Thus, f is independent of y , say $f(x, y) = F(x)$. The equation $Lf = 0$ then gives

$$\sin(y)F'(x) = 0,$$

hence F is constant. Thus, the kernel of L consists only of constants.

One may also check that L has compact resolvent by showing that the embedding $(\operatorname{Dom}(L), \|\cdot\| + \|L\cdot\|) \hookrightarrow L^2(\mathbb{T}^2)$ is compact. See Appendix A.4 for details. As a result, conditions for Proposition 6 are satisfied and

$$\gamma = \inf_{f \in \operatorname{Dom}(L) \cap L_0^2(\mu) \setminus \{0\}} \frac{\|Lf\|}{\|f\|} > 0.$$

Consequently, Corollary 1 yields

$$\operatorname{Var}_\mu(\mu_t g) \leq \frac{4}{\gamma t} \|g\|^2, \quad g \in L_0^2(\mu).$$

Although the spectrum of L is not explicit in this example, we take advantage of the standard tools for hypoelliptic differential operators and the compact-resolvent criterion in Proposition 6 to show that $\gamma > 0$. Thus, even though the symmetric part ∂_y^2 has no spectral gap on $L_0^2(\mu)$, the coupling created by the shear drift $\sin(y)\partial_x$ is enough to give a uniform stationary empirical-variance bound.

More generally, the compact-resolvent assumption of Proposition 6 is natural for hypoelliptic operators on compact manifolds. Hörmander's bracket condition can be checked to prove hypoellipticity for operators of the form $L = X_0 + \sum_i X_i^2$, and subelliptic estimates often imply compactness of the embedding of $\operatorname{Dom}(L)$ into L^2 .

APPENDIX A. OMITTED DETAILS

A.1. The Poisson-equation formula for asymptotic variance. We record a derivation of the asymptotic variance formula (7). This formula is known in the literature, but often with extra hypotheses, such as when a central limit theorem is sought in addition to a limiting second-moment. Thus, we provide a clear statement and proof in accordance with our operator-theoretic perspective in this paper.

Lemma 5 (Poisson equation formula for asymptotic variance). *Consider a strongly continuous Markov semigroup $(P_t)_{t \geq 0}$ on a measurable space with an invariant probability measure μ . Let $K \subset L_0^2(\mu)$ be a closed P_t -invariant subspace, and let L_K denote the generator of the restricted semigroup $(P_t|_K)_{t \geq 0}$. Fix $g \in K$. Suppose there exists $f \in \text{Dom}(L_K)$ such that*

$$-L_K f = g.$$

Then the asymptotic variance of g exists and is given by

$$\sigma_g^2 := \lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t g) = 2 \text{Re}\langle g, f \rangle.$$

In particular, if the local gap defined in (19) satisfies $\gamma_K > 0$, then

$$\sigma_g^2 = 2 \text{Re}\langle g, -L_K^{-1} g \rangle.$$

Proof. To simplify the notation, denote by $(P_t)_{t \geq 0}$ the semigroup restricted to K , and let A_t be the corresponding restricted Cesàro-mean operator, defined in (13). By (16),

$$\text{Var}_\mu(\mu_t g) = \frac{2}{t^2} \text{Re} \int_0^t (t-s) \langle g, P_s g \rangle ds.$$

Since $f \in \text{Dom}(L_K)$ and $-L_K f = g$, the map $s \mapsto P_s f$ is differentiable in $L^2(\mu)$, and

$$\frac{d}{ds} P_s f = P_s L_K f = -P_s g.$$

Therefore

$$\frac{d}{ds} \langle g, P_s f \rangle = \langle g, -P_s g \rangle = -\langle g, P_s g \rangle.$$

Thus

$$\int_0^t (t-s) \langle g, P_s g \rangle ds = - \int_0^t (t-s) \frac{d}{ds} \langle g, P_s f \rangle ds = t \langle g, f \rangle - \int_0^t \langle g, P_s f \rangle ds.$$

Multiplying by $2/t$ and using the definition of A_t , we obtain

$$t \text{Var}_\mu(\mu_t g) = 2 \text{Re}\langle g, f \rangle - 2 \text{Re}\langle g, A_t f \rangle. \quad (46)$$

It remains to show that as $t \rightarrow \infty$,

$$\langle g, A_t f \rangle \rightarrow 0.$$

By Theorem V.4.5 and Example V.4.7 in [EN00], $(P_t)_{t \geq 0}$ is mean ergodic, so the strong limit

$$\Pi h := \lim_{t \rightarrow \infty} A_t h$$

exists for all $h \in K$. By Lemma V.4.4 in [EN00], Π is the mean-ergodic orthogonal projection onto $\text{Ran } \Pi = \ker L_K$. Since $g = -L_K f$, by (14) we have

$$\|A_t g\| = \frac{\|f - P_t f\|}{t} \leq \frac{2}{t} \|f\| \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

Therefore, $\Pi g = 0$, hence $g \perp \text{Ran } \Pi$. In particular,

$$\langle g, A_t f \rangle \rightarrow \langle g, \Pi f \rangle = 0.$$

Returning to (46), we conclude that

$$\sigma_g^2 = \lim_{t \rightarrow \infty} t \text{Var}_\mu(\mu_t g) = 2 \text{Re} \langle g, f \rangle.$$

Finally, when $\gamma_K > 0$, L_K is invertible by Lemma 2, so we may take $f = -L_K^{-1} g$. $\square$

A.2. Eigenfunctions of the rotating Ornstein–Uhlenbeck generator. In polar coordinates, the generator can be decomposed as

$$L = S + \alpha A, \quad \text{where } S = \Delta - x \cdot \nabla = \partial_{rr} + \left(\frac{1}{r} - r\right) \partial_r + \frac{1}{r^2} \partial_{\theta\theta}, \quad A = \partial_\theta$$

We look for eigenfunctions of the form

$$\psi(r, \theta) = R(r) e^{im\theta}, \quad m \in \mathbb{Z}.$$

Then $\partial_\theta \psi = im\psi$, and $S\psi = -\lambda\psi$ becomes

$$R''(r) + \left(\frac{1}{r} - r\right) R'(r) - \frac{m^2}{r^2} R(r) = -\lambda R(r).$$

Let $a = |m|$, $\rho = r^2/2$, and set

$$R(r) = r^a Q(\rho).$$

A direct calculation gives

$$\rho Q''(\rho) + (a + 1 - \rho) Q'(\rho) + \frac{\lambda - a}{2} Q(\rho) = 0. \quad (47)$$

This is the generalized Laguerre equation. Polynomial solutions occur when

$$\frac{\lambda - a}{2} = k, \quad k = 0, 1, 2, \dots,$$

so by, say, p.781 of [AS72], the equation (47) is solved by

$$\lambda = 2k + a, \quad Q(\rho) = L_k^{(a)}(\rho).$$

Thus

$$\psi_{k,m}(r, \theta) = c_{k,m} r^{|m|} L_k^{(|m|)}(r^2/2) e^{im\theta}$$

satisfies

$$S\psi_{k,m} = -(2k + |m|)\psi_{k,m}, \quad \partial_\theta \psi_{k,m} = im\psi_{k,m}.$$

The normalization constant $c_{k,m}$ can be computed from the orthogonality relation

$$\int_0^\infty \rho^a e^{-\rho} L_k^{(a)}(\rho) L_\ell^{(a)}(\rho) d\rho = \frac{\Gamma(k+a+1)}{k!} \mathbf{1}_{k=\ell}.$$

Since $d\mu = (2\pi)^{-1} e^{-r^2/2} r dr d\theta$, this gives

$$c_{k,m} = \left(\frac{k!}{2^{|m|} \Gamma(k + |m| + 1)} \right)^{1/2}.$$

A.3. Spectrum of the Jacobi operator for the reflected birth-death chain. We consider the jump process in Example 7 and compute its spectral gap. Recall that the generator of this process is given by

$$\begin{aligned} (Lf)(0) &= \lambda(f(1) - f(0)), \\ (Lf)(n) &= \lambda(f(n+1) - f(n)) + \mu(f(n-1) - f(n)), \quad n \geq 1. \end{aligned}$$

The invariant measure is given by $\pi(n) = (1 - \rho)\rho^n$, $\rho = \frac{\lambda}{\mu} < 1$. Conjugating $-L$ from $L^2(\pi)$ to $\ell^2(\mathbb{N})$ by the spectrum-preserving unitary isometry

$$(Uf)(n) = \sqrt{\pi(n)} f(n)$$

gives the self-adjoint operator $H = U(-L)U^{-1}$ on $\ell^2(\mathbb{N})$ that satisfies

$$(Hg)(0) = \lambda g(0) - \sqrt{\lambda\mu} g(1),$$

and, for $n \geq 1$,

$$(Hg)(n) = (\lambda + \mu)g(n) - \sqrt{\lambda\mu} g(n+1) - \sqrt{\lambda\mu} g(n-1).$$

The vector $(\sqrt{\pi(n)})_{n \geq 0}$ is an eigenvector with eigenvalue 0. Set

$$c := \sqrt{\lambda\mu}.$$

Away from the boundary, H has constant Jacobi coefficients: diagonal coefficient $\lambda + \mu$ and off-diagonal coefficient $-c$. By Theorem 3.19 in [Tes00]

$$\sigma_{\text{ess}}(H) = [\lambda + \mu - 2c, \lambda + \mu + 2c]. \quad (48)$$

Therefore, to prove that the spectral gap is $\gamma = (\sqrt{\mu} - \sqrt{\lambda})^2 = \lambda + \mu - 2c$, it remains to check that there are no eigenvalues in the interval $(0, \lambda + \mu - 2c)$. Suppose that $\theta < \lambda + \mu - 2c$ is an eigenvalue of H , with eigenvector $g \in \ell^2(\mathbb{N})$. For $n \geq 1$, the eigenvalue equation is

$$(\lambda + \mu - \theta)g(n) - c g(n+1) - c g(n-1) = 0.$$

This is a system of linear recurrence equations, and the solution is determined by $g(0)$ and $g(1)$. With the ansatz $g(n) = r^n$, we get

$$c \left(r + \frac{1}{r} \right) = \lambda + \mu - \theta. \quad (49)$$

Since $\theta < \lambda + \mu - 2c$, this equation has one root $r \in (0, 1)$ and one root $r^{-1} > 1$, and we deduce that all solutions must be of the form $g(n) = Ar^n + Br^{-n}$. The condition $g \in \ell^2(\mathbb{N})$ forces $B = 0$, so for $n \geq 1$,

$$g(n) = Ar^n.$$

We now use the bulk equation at $n = 1$:

$$(\lambda + \mu - \theta)g(1) - cg(2) - cg(0) = 0.$$

Substituting $g(1) = Ar$, $g(2) = Ar^2$, and (49), we obtain

$$c(r + r^{-1})Ar - cAr^2 - cg(0) = 0,$$

hence $g(0) = A$. Now use the boundary equation at $n = 0$:

$$\lambda g(0) - c g(1) = \theta g(0).$$

Since $g(1) = rg(0)$, this gives

$$\theta = \lambda - cr.$$

On the other hand, (49) gives

$$\theta = \lambda + \mu - c \left(r + \frac{1}{r} \right).$$

Comparing the two expressions for θ , we obtain

$$r = \frac{c}{\mu} = \sqrt{\frac{\lambda}{\mu}}.$$

Substituting this into $\theta = \lambda - cr$ yields

$$\theta = \lambda - \sqrt{\lambda\mu} \sqrt{\frac{\lambda}{\mu}} = 0.$$

Therefore, the only eigenvalue of H below $\lambda + \mu - 2\sqrt{\lambda\mu}$ is 0. Together with (48) and the fact that $\sigma(H) = \sigma(-L)$, this shows that the spectral gap of L is

$$\gamma = (\sqrt{\mu} - \sqrt{\lambda})^2.$$

A.4. Compactness for the hypoelliptic shear on the torus. We prove here that for the Markov diffusion generator

$$L = \partial_y^2 + \sin(y)\partial_x$$

on $L^2(\mathbb{T}^2)$, the embedding

$$(\text{Dom}(L), \|\cdot\| + \|L\cdot\|) \hookrightarrow L^2(\mathbb{T}^2)$$

is compact. For $f \in \text{Dom}(L)$, write

$$f(x, y) = \sum_{m \in \mathbb{Z}} e^{imx} \phi_m(y).$$

Since the coefficients of L are independent of x , the x -Fourier modes are invariant:

$$L(e^{imx} \phi(y)) = e^{imx} L_m \phi(y), \quad L_m := \partial_y^2 + im \sin(y).$$

By Parseval,

$$\|f\|^2 + \|Lf\|^2 = \sum_{m \in \mathbb{Z}} \left(\|\phi_m\|_{L_y^2}^2 + \|L_m \phi_m\|_{L_y^2}^2 \right).$$

For brevity, set

$$E_m := \|\phi_m\|_{L_y^2}^2 + \|L_m \phi_m\|_{L_y^2}^2.$$

We have the following estimates, whose proof we defer for now.

Lemma 6. *With the above notations, we have for some constant C that*

$$\begin{aligned} \|\phi_m\|_{H_y^1}^2 &\leq C E_m \\ \|\phi_m\|_{L_y^2}^2 &\leq C |m|^{-1/3} E_m, \quad m \neq 0. \end{aligned}$$

Let (f_n) be a sequence in $\text{Dom}(L)$ satisfying

$$\|f_n\| + \|Lf_n\| \leq 1.$$

Write

$$f_n(x, y) = \sum_{m \in \mathbb{Z}} e^{imx} \phi_{n,m}(y),$$

and define $E_{n,m}$ as above. Then

$$\sum_m E_{n,m} \leq C$$

uniformly in n . The high- x -frequency estimate in Lemma 6 gives, for $M \geq 1$,

$$\sum_{|m| > M} \|\phi_{n,m}\|_{L_y^2}^2 \leq CM^{-1/3} \sum_{|m| > M} E_{n,m} \leq CM^{-1/3}, \quad (50)$$

uniformly in n . For each fixed m , the estimate

$$\|\phi_{n,m}\|_{H_y^1}^2 \leq C E_{n,m} \leq C$$

shows that $(\phi_{n,m})_n$ is bounded in $H^1(\mathbb{T}_y)$. By the Rellich compactness theorem, $H^1(\mathbb{T}_y) \hookrightarrow L^2(\mathbb{T}_y)$ compactly. Therefore, by a diagonal argument, we can extract a subsequence, reindexed by n , such that $(\phi_{n,m})_n$ converges in L_y^2 for every m . Using the tail estimate (50), we see that $(f_n)_n$ is Cauchy, hence convergent in $L^2(\mathbb{T}^2)$. Thus, the graph-norm unit ball

$$\{f : \|f\| + \|Lf\| \leq 1\}$$

is relatively compact in $L^2(\mathbb{T}^2)$. Equivalently, we have a compact embedding

$$(\text{Dom}(L), \|\cdot\| + \|L\cdot\|) \hookrightarrow L^2(\mathbb{T}^2).$$

Proof of Lemma 6. Since smooth functions are dense in $\text{Dom}(L)$, we can assume ϕ_m to be smooth. Then

$$\text{Re}\langle \phi_m, L_m \phi_m \rangle = \text{Re}\langle \phi_m, \phi_m'' \rangle = -\|\phi_m'\|_{L_y^2}^2.$$

Therefore,

$$\|\phi_m'\|_{L_y^2}^2 \leq |\langle \phi_m, L_m \phi_m \rangle| \leq \|\phi_m\|_{L_y^2} \|L_m \phi_m\|_{L_y^2} \leq \frac{1}{2} E_m.$$

In particular,

$$\|\phi_m\|_{H_y^1}^2 \leq C E_m.$$

Next, for $|m| \geq 1$, we estimate $\sin(y)\phi_m$. Since

$$L_m \phi_m = \phi_m'' + im \sin(y) \phi_m,$$

we have

$$im \|\sin(y)\phi_m\|_{L_y^2}^2 = \langle \sin(y)\phi_m, L_m \phi_m \rangle - \langle \sin(y)\phi_m, \phi_m'' \rangle.$$

Taking imaginary parts gives

$$|m| \|\sin(y)\phi_m\|_{L_y^2}^2 \leq |\langle \sin(y)\phi_m, L_m \phi_m \rangle| + |\text{Im}\langle \sin(y)\phi_m, \phi_m'' \rangle|.$$

The first term is bounded by

$$\|\phi_m\|_{L_y^2} \|L_m \phi_m\|_{L_y^2} \leq \frac{1}{2} E_m.$$

For the second term, integrate by parts:

$$\langle \sin(y)\phi_m, \phi_m'' \rangle = - \int_{\mathbb{T}} \cos(y) \overline{\phi_m(y)} \phi_m'(y) dy - \int_{\mathbb{T}} \sin(y) |\phi_m'(y)|^2 dy.$$

The second integral is real, and hence

$$|\text{Im}\langle \sin(y)\phi_m, \phi_m'' \rangle| \leq \|\phi_m\|_{L_y^2} \|\phi_m'\|_{L_y^2} \leq C E_m.$$

Thus,

$$|m| \|\sin(y)\phi_m\|_{L_y^2}^2 \leq C E_m.$$

We now prove that high x -frequencies have small L^2 -mass. Let

$$Z_\delta := \{y \in \mathbb{T} : |\sin y| \leq \delta\}.$$

For $0 < \delta \leq 1/2$, $|Z_\delta| \leq C\delta$. Since $H^1(\mathbb{T}) \hookrightarrow L^\infty(\mathbb{T})$ continuously,

$$\int_{Z_\delta} |\phi_m(y)|^2 dy \leq |Z_\delta| \|\phi_m\|_{L_y^\infty}^2 \leq C\delta \|\phi_m\|_{H_y^1}^2 \leq C\delta E_m.$$

On the complement Z_δ^c , we have $|\sin y| > \delta$, and so

$$\int_{Z_\delta^c} |\phi_m(y)|^2 dy \leq \delta^{-2} \|\sin(y)\phi_m\|_{L_y^2}^2 \leq C\delta^{-2} |m|^{-1} E_m.$$

Choosing $\delta = |m|^{-1/3}$ gives, for $|m| \geq 1$,

$$\|\phi_m\|_{L_y^2}^2 \leq C|m|^{-1/3} E_m.$$

□

ACKNOWLEDGEMENTS

I am grateful to Sourav Chatterjee for the suggestion to look for extensions of his results, and to Amir Dembo for encouraging me in the research process. I thank both of them and Persi Diaconis for helpful feedback.

REFERENCES

- [AF14] David J. Aldous and James A. Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2014. URL: <https://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [AS72] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Vol. 55. Tenth Printing, with corrections. Dover Publications, New York, 1972. URL: https://personal.math.ubc.ca/~cbm/aands/abramowitz_and_stegun.pdf.
- [BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Vol. 348. Springer, 2014.
- [Bha82] Rabi N Bhattacharya. “On the functional central limit theorem and the law of the iterated logarithm for Markov processes”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 60.2 (1982), pp. 185–201. URL: <https://doi.org/10.1007/BF00531822>.
- [CCG12] Patrick Cattiaux, Djalil Chafai, and Arnaud Guillin. “Central limit theorems for additive functionals of ergodic Markov diffusions processes”. In: *ALEA, Latin American Journal of Probability and Mathematical Statistics* 9.2 (2012), pp. 337–382. URL: <https://alea.impa.br/articles/v9/09-15.pdf>.
- [Cha25] Sourav Chatterjee. “Spectral gap of nonreversible Markov chains”. In: *The Annals of Applied Probability* 35.4 (2025), pp. 2644–2677. URL: <https://doi.org/10.1214/25-AAP2183>.
- [CL14] Yong Chen and Yong Liu. “On the eigenfunctions of the complex Ornstein–Uhlenbeck operators”. In: *Kyoto Journal of Mathematics* 54.3 (Jan. 2014). URL: <http://dx.doi.org/10.1215/21562261-2693451>.
- [CLP99] Fang Chen, László Lovász, and Igor Pak. “Lifting Markov chains to speed up mixing”. In: *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*. STOC ’99. Atlanta, Georgia, USA: Association for Computing Machinery, 1999, pp. 275–281. URL: <https://doi.org/10.1145/301250.301315>.
- [DHN00] Persi Diaconis, Susan Holmes, and Radford M. Neal. “Analysis of a nonreversible Markov chain sampler”. In: *The Annals of Applied Probability* 10.3 (2000), pp. 726–752. URL: <https://doi.org/10.1214/aoap/1019487508>.
- [Din98] I. H. Dinwoodie. “Expectations for Nonreversible Markov Chains”. In: *Journal of Mathematical Analysis and Applications* 220.2 (1998), pp. 585–596. URL: <https://www.sciencedirect.com/science/article/pii/S0022247X97958509>.
- [DLP16] A. B. Duncan, T. Lelièvre, and G. A. Pavliotis. “Variance Reduction Using Nonreversible Langevin Samplers”. In: *Journal of Statistical Physics* 163.3 (Mar. 2016), pp. 457–491. URL: <http://dx.doi.org/10.1007/s10955-016-1491-2>.
- [DM13] Persi Diaconis and Laurent Miclo. “On the spectral analysis of second-order Markov chains”. In: *Annales de la Faculté des sciences de Toulouse : Mathématiques* Ser. 6, 22.3 (2013), pp. 573–621. URL: <https://www.numdam.org/articles/10.5802/afst.1383/>.
- [DPU15] George Deligiannidis, Magda Peligrad, and Sergey Utev. “Asymptotic variance of stationary reversible and normal Markov processes”. In: *Electronic Journal of Probability* 20 (2015), pp. 1–26. URL: <https://doi.org/10.1214/EJP.v20-3183>.
- [DS93] Persi Diaconis and Laurent Saloff-Coste. “Comparison Theorems for Reversible Markov Chains”. In: *The Annals of Applied Probability* 3.3 (1993), pp. 696–730. URL: <https://doi.org/10.1214/aoap/1177005359>.
- [Ebe] Andreas Eberle. *Markov Processes*. Unpublished lecture notes for the course “Markov Processes”, University of Bonn, Winter term 2025/26. URL: <https://uni-bonn.sciebo.de/s/kzTUFFf5FrWGAay?dir=/&editing=false&openfile=true>.
- [EL24] Andreas Eberle and Francis Lörler. *Non-reversible lifts of reversible diffusion processes and relaxation times*. 2024. arXiv: 2402.05041 [math.PR].

- [EN00] Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Vol. 194. Graduate Texts in Mathematics. Springer, 2000.
- [Fil91] James Allen Fill. “Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process”. In: *The Annals of Applied Probability* 1.1 (1991), pp. 62–87. URL: <https://doi.org/10.1214/aoap/1177005981>.
- [FOT10] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. *Dirichlet Forms and Symmetric Markov Processes*. 2nd. Vol. 19. de Gruyter Studies in Mathematics. Walter de Gruyter, 2010.
- [HHS05] Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. “Accelerating Diffusions”. In: *The Annals of Applied Probability* 15.2 (2005), pp. 1433–1444. URL: <http://www.jstor.org/stable/30038359>.
- [HHS93] Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. “Accelerating Gaussian diffusions”. In: *The Annals of Applied Probability* 3.3 (1993), pp. 897–913. URL: <https://doi.org/10.1214/aoap/1177005371>.
- [HJ01] James P. Hobert and Galin L. Jones. “Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo”. In: *Statistical Science* 16.4 (2001), pp. 312–334. URL: <https://doi.org/10.1214/ss/1015346317>.
- [HL25] De Huang and Xiangyuan Li. *Bernstein-type Inequalities for Markov Chains and Markov Processes: A Simple and Robust Proof*. 2025. arXiv: 2408.04930 [math.PR].
- [HM11] Martin Hairer and Jonathan C. Mattingly. “Yet Another Look at Harris’ Ergodic Theorem for Markov Chains”. In: *Seminar on Stochastic Analysis, Random Fields and Applications VI*. Ed. by Robert Dalang, Marco Dozzi, and Francesco Russo. Springer Basel, 2011, pp. 109–117.
- [HNW15] Chii-Ruey Hwang, Raoul Normand, and Sheng-Jhih Wu. “Variance reduction for diffusions”. In: *Stochastic Processes and their Applications* 125.9 (2015), pp. 3522–3540. URL: <https://www.sciencedirect.com/science/article/pii/S030441491500085X>.
- [Hol05] Hajo Holzmann. “Martingale approximations for continuous-time and discrete-time stationary Markov processes”. In: *Stochastic Processes and their Applications* 115.9 (2005), pp. 1518–1529. URL: <https://www.sciencedirect.com/science/article/pii/S0304414905000463>.
- [Hör67] Lars Hörmander. “Hypoelliptic second order differential equations”. In: *Acta Mathematica* 119 (1967), pp. 147–171. URL: <https://doi.org/10.1007/BF02392081>.
- [JO10] Aldéric Joulin and Yann Ollivier. “Curvature, concentration and error estimates for Markov chain Monte Carlo”. In: *The Annals of Probability* 38.6 (Nov. 2010). URL: <http://dx.doi.org/10.1214/10-AOP541>.
- [Kat95] Tosio Kato. *Perturbation Theory for Linear Operators*. Reprint of the 1980 edition. Classics in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.
- [KLO12] Tomasz Komorowski, Claudio Landim, and Stefano Olla. *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation*. Springer, 2012.
- [KM03] I. Kontoyiannis and S. P. Meyn. “Spectral theory and limit theorems for geometrically ergodic Markov processes”. In: *The Annals of Applied Probability* 13.1 (2003), pp. 304–362. URL: <https://doi.org/10.1214/aoap/1042765670>.
- [KM11] Ioannis Kontoyiannis and Sean P. Meyn. “Geometric ergodicity and the spectral gap of non-reversible Markov chains”. In: *Probability Theory and Related Fields* 151.1 (2011), pp. 1–35. URL: <https://link.springer.com/article/10.1007/s00440-011-0373-4>.
- [KV86] Claude Kipnis and Srinivasa R. S. Varadhan. “Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions”. In: *Communications in Mathematical Physics* 104.1 (1986), pp. 1–19. URL: <https://doi.org/10.1007/BF01210789>.
- [Lez01] Pascal Lezaud. “Chernoff and Berry-Esséen inequalities for Markov processes”. In: *ESAIM: Probability and Statistics* 5 (2001), pp. 183–201. URL: https://www.numdam.org/item/PS_2001__5__183_0/.

- [LNP13] T. Lelièvre, F. Nier, and G. A. Pavliotis. “Optimal Non-reversible Linear Drift for the Convergence to Equilibrium of a Diffusion”. In: *Journal of Statistical Physics* 152.2 (2013), pp. 237–274. URL: <http://dx.doi.org/10.1007/s10955-013-0769-x>.
- [LP17] David Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2017.
- [LS88] Gregory Lawler and Alan Sokal. “Bounds on the L^2 Spectrum for Markov Chains and Markov Processes: A Generalization of Cheeger’s Inequality”. In: *Transactions of the American Mathematical Society* 309.2 (1988), pp. 557–580. URL: <http://www.jstor.org/stable/2000925>.
- [MSH02] J.C. Mattingly, A.M. Stuart, and D.J. Higham. “Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise”. In: *Stochastic Processes and their Applications* 101.2 (2002), pp. 185–232. URL: <https://www.sciencedirect.com/science/article/pii/S0304414902001503>.
- [ODo14] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.
- [Pau18] Daniel Paulin. *Concentration inequalities for Markov chains by Marton couplings and spectral methods*. 2018. arXiv: [1212.2015](https://arxiv.org/abs/1212.2015) [math.PR].
- [RR04] Gareth O. Roberts and Jeffrey S. Rosenthal. “General state space Markov chains and MCMC algorithms”. In: *Probability Surveys* 1 (2004), pp. 20–71. URL: <https://doi.org/10.1214/154957804100000024>.
- [RS15] Luc Rey-Bellet and Konstantinos Spiliopoulos. “Irreversible Langevin samplers and variance reduction: a large deviations approach”. In: *Nonlinearity* 28.7 (May 2015), pp. 2081–2103. URL: <http://dx.doi.org/10.1088/0951-7715/28/7/2081>.
- [RS18] Kavita Ramanan and Aaron Smith. “Bounds on Lifting Continuous-State Markov Chains to Speed Up Mixing”. In: *Journal of Theoretical Probability* 31.3 (2018), pp. 1647–1678. URL: <https://doi.org/10.1007/s10959-017-0745-5>.
- [RS76] Linda Preiss Rothschild and E. M. Stein. “Hypoelliptic differential operators and nilpotent groups”. In: *Acta Mathematica* 137 (1976), pp. 247–320. URL: <https://doi.org/10.1007/BF02392419>.
- [SV05] Sunder Sethuraman and Srinivasa Varadhan. “A Martingale Proof of Dobrushin’s Theorem for Non-Homogeneous Markov Chains”. In: *Electronic Journal of Probability* 10 (2005), pp. 1221–1235. URL: <https://doi.org/10.1214/EJP.v10-283>.
- [Tes00] Gerald Teschl. *Jacobi Operators and Completely Integrable Nonlinear Lattices*. Vol. 72. Mathematical Surveys and Monographs. American Mathematical Society, 2000.