

Accelerating Constrained Sampling: A Large Deviations Approach

Yingli Wang

NATURESKY1994@GMAIL.COM

School of Mathematics

Shanghai University of Finance and Economics

Shanghai, People's Republic of China

FinTech Thrust

Hong Kong University of Science and Technology (Guangzhou)

Guangzhou, Guangdong Province, People's Republic of China

Changwei Tu

CTU570@CONNECT.HKUST-GZ.EDU.CN

FinTech Thrust

Hong Kong University of Science and Technology (Guangzhou)

Guangzhou, Guangdong Province, People's Republic of China

Xiaoyu Wang

XIAOYUWANG@HKUST-GZ.EDU.CN

FinTech Thrust

Hong Kong University of Science and Technology (Guangzhou)

Guangzhou, Guangdong Province, People's Republic of China

Lingjiong Zhu

ZHU@MATH.FSU.EDU

Department of Mathematics

Florida State University

Tallahassee, Florida, United States of America

Editor: Jianfeng Lu

Abstract

The problem of sampling a target probability distribution on a constrained domain arises in many applications including machine learning. For constrained sampling, various Langevin algorithms such as projected Langevin Monte Carlo (PLMC), based on the discretization of reflected Langevin dynamics (RLD) and more generally skew-reflected non-reversible Langevin Monte Carlo (SRNLMC), based on the discretization of skew-reflected non-reversible Langevin dynamics (SRNLD), have been proposed and studied in the literature. This work focuses on the long-time behavior of SRNLD, where a skew-symmetric matrix is added to RLD. Although acceleration for SRNLD has been studied, it is not clear how one should design the skew-symmetric matrix in the dynamics to achieve good performance in practice. We establish a large deviation principle (LDP) for the empirical measure of SRNLD when the skew-symmetric matrix is chosen such that its product with the outward unit normal vector field on the boundary is zero. By explicitly characterizing the rate functions, we show that this choice of the skew-symmetric matrix accelerates the convergence to the target distribution compared to RLD and reduces the asymptotic variance. Numerical experiments for SRNLMC based on the proposed skew-symmetric matrix show superior performance, which validate the theoretical findings from the large deviations theory.

Keywords: Constrained Sampling, Non-reversible Langevin Dynamics, Large Deviations Theory, Variance Reduction

1. Introduction

We consider the problem of sampling a distribution μ on a constrained domain $K \subset \mathbb{R}^d$ with probability density function

$$\mu(x) \propto \exp(-f(x)), \quad x \in K, \quad (1.1)$$

for a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. The sampling problem for both constrained domain $K \subset \mathbb{R}^d$ and unconstrained domain $K = \mathbb{R}^d$ is fundamental in many applications, such as Bayesian statistical inference, Bayesian formulations of inverse problems, and Bayesian classification and regression tasks in machine learning (Gelman et al., 1995; Stuart, 2010; Andrieu et al., 2003; Teh et al., 2016; Gürbüzbalaban et al., 2021; Gürbüzbalaban et al., 2024a,b).

In the literature, Bubeck et al. (2015, 2018) proposed and studied the *projected Langevin Monte Carlo* (PLMC) algorithm for constrained sampling, which projects each unconstrained Langevin step back to the constraint set K :

$$x_{k+1} = \mathcal{P}_K \left(x_k - \eta \nabla f(x_k) + \sqrt{2\eta} \xi_{k+1} \right), \quad (1.2)$$

where $\eta > 0$ is the stepsize and $\{\xi_k\}_{k \geq 1}$ are i.i.d. Gaussian random vectors with $\xi_k \sim \mathcal{N}(0, I_d)$, and the mapping \mathcal{P}_K denotes the (Euclidean) metric projection onto K , defined for any $x \in \mathbb{R}^d$ by

$$\mathcal{P}_K(x) := \arg \min_{y \in K} \|x - y\|_2. \quad (1.3)$$

In particular, when K is nonempty, closed, and convex, $\mathcal{P}_K(x)$ is well-defined and single-valued.

The dynamics (1.2) is motivated by the *reflected Langevin dynamics* (RLD), i.e. the continuous-time overdamped Langevin SDE with reflection at the boundary:

$$dX_t = -\nabla f(X_t) dt + \sqrt{2} dW_t - \mathbf{n}(X_t) Z(dt), \quad (1.4)$$

where W_t is a standard d -dimensional Brownian motion, $\mathbf{n}(x)$ is the outward unit normal vector for $x \in \partial K$ (and can be chosen arbitrarily off ∂K), and $\int_0^t \mathbf{n}(X_s) Z(ds)$ is a bounded-variation reflection term that enforces the constraint $X_t \in K$ for all $t \geq 0$ whenever $X_0 \in K$. The measure $Z(dt)$ is nonnegative with $Z([0, t]) < \infty$ for each t , and it charges only boundary times, i.e. $\text{supp}(Z) \subseteq \{t \geq 0 : X_t \in \partial K\}$.

The reflected Langevin dynamics (1.4) can be viewed as a Skorokhod problem on K : one first follows the unconstrained Langevin increment, and whenever the trajectory hits the boundary, a bounded-variation correction is added so that $X_t \in K$ for all $t \geq 0$. Geometrically, this correction acts only at boundary times (i.e., the measure $Z(dt)$ is supported on $\{t : X_t \in \partial K\}$) and pushes the trajectory inward along $-\mathbf{n}$. In smooth domains, the reflection term is the *minimal* correction that prevents leaving K , in contrast to discrete-time schemes that reproject the iterate after each time step.

Since the seminal works (Bubeck et al., 2015, 2018) where the density μ is assumed to be log-concave, there have been many studies proposing various Langevin algorithms for constrained sampling. Lamperski (2021) considers the *projected stochastic gradient Langevin dynamics* in the setting of non-convex smooth Lipschitz f on a convex body where the gradient noise is assumed to have finite variance with a uniform sub-Gaussian structure;

see also Zheng and Lamperski (2022). *Proximal Langevin Monte Carlo* is proposed in Brosse et al. (2017) for constrained sampling; see also Salim and Richtárik (2020) for a study on the proximal stochastic gradient Langevin algorithm from a primal-dual perspective. *Mirror descent-based Langevin algorithms* (see e.g. Hsieh et al. (2018); Chewi et al. (2020); Zhang et al. (2020); Li et al. (2022); Ahn and Chewi (2021)) can also be used for constrained sampling, that was first proposed in Hsieh et al. (2018), inspired by the classical mirror descent in optimization. More recently, inspired by the penalty method in the optimization literature, *penalized Langevin Monte Carlo algorithms* are proposed and studied in Gürbüzbalaban et al. (2024a), where the objective f can be non-convex in general. In addition, constrained non-convex exploration combined with replica-exchange Langevin dynamics was proposed and studied in Zheng et al. (2024).

In the literature, it is known that in the setting of unconstrained sampling, by *breaking reversibility*, a *non-reversible* Langevin dynamics, where an *skew-symmetric* matrix is added, can converge to the target distribution faster on the Euclidean space \mathbb{R}^d . This phenomenon is well understood through non-asymptotic convergence analysis; see e.g. Hwang et al. (1993, 2005); Rey-Bellet and Spiliopoulos (2015); Hu et al. (2020); Gao et al. (2020). Motivated by this, in the setting of constrained sampling, non-reversible Langevin algorithms with skew-reflected boundary was proposed and studied in Du et al. (2025), that can accelerate the convergence of the (reversible) projected Langevin Monte Carlo, based on the discretization of the *skew-reflected non-reversible Langevin dynamics* (SRNLD):

$$dX_t = -(I + J(X_t))\nabla f(X_t)dt + \sqrt{2}dW_t - \mathbf{n}^J(X_t)L(dt), \quad (1.5)$$

where W_t is a standard d -dimensional Brownian motion; for every x , $J(x)$ is a skew-symmetric matrix, i.e. $J(x) = -(J(x))^\top$, and moreover, $\|J\|_\infty := \sup_{x \in K} \|J(x)\| < \infty$ and $\nabla \cdot J = 0$. The term $\mathbf{n}^J(X_t)L(dt)$ ensures that $X_t \in K$ for every t given that $X_0 \in K$. In particular, $\int_0^t \mathbf{n}^J(X_s)L(ds)$ is a bounded variation *skew-reflection* process and the measure $L(dt)$ is such that $L([0, t])$ is finite, $L(dt)$ is supported on $\{t : X_t \in \partial K\}$, where the skew-reflection is defined through the following *skew unit normal vector*:

$$\mathbf{n}^J(X_t) := \frac{(I + J(X_t))\mathbf{n}(X_t)}{\sqrt{\|\mathbf{n}(X_t)\|_2^2 + \|J(X_t)\mathbf{n}(X_t)\|_2^2}}, \quad (1.6)$$

where $\mathbf{n}(x)$ is the *outward* unit normal vector for $x \in \partial K$ (and can be chosen arbitrarily off ∂K), and $\mathcal{N}_K(x)$ denotes the (Euclidean) normal cone of K at x , defined by

$$\mathcal{N}_K(x) := \left\{ v \in \mathbb{R}^d : \langle v, y - x \rangle \leq 0, \text{ for any } y \in K \right\}, \quad x \in K.$$

(For $x \in \text{int}(K)$, $\mathcal{N}_K(x) := \{0\}$.)

We assume the skew-matrix J satisfies the condition that $\langle \mathbf{n}^J, \mathbf{n} \rangle \geq \delta_0 > 0$ for some positive constant δ_0 . To implement SRNLD (1.5), *skew-reflected non-reversible Langevin Monte Carlo* (SRNLMC) algorithm was proposed and studied in Du et al. (2025) that is based on the discretization of (1.5):

$$x_{k+1} = \mathcal{P}_K^J \left(x_k - \eta(I + J(x_k))\nabla f(x_k) + \sqrt{2\eta}\xi_{k+1} \right), \quad (1.7)$$

where ξ_k are i.i.d. Gaussian random vectors $\mathcal{N}(0, I_d)$, and \mathcal{P}_K^J is the *skew-projection*:

$$\mathcal{P}_K^J(x) := \arg \max_{y \in K} \langle y - x, \mathbf{n}^J(\mathcal{P}_K(x)) \rangle, \quad (1.8)$$

where \mathcal{P}_K is the Euclidean projection defined in (1.3). The mapping \mathcal{P}_K^J in (1.8) is an *oblique* (or skew) projection associated with the skew-reflection direction \mathbf{n}^J . Unlike the Euclidean metric projection \mathcal{P}_K in (1.3), which returns the closest point in K to a given x , the skew-projection corrects infeasible points by moving them back to K *along the prescribed direction* \mathbf{n}^J (computed at the Euclidean projection point $\mathcal{P}_K(x)$). In two dimensions, one may visualize $\mathcal{P}_K^J(x)$ as the boundary point obtained by transporting x in the direction $-\mathbf{n}^J(\mathcal{P}_K(x))$ until it re-enters the feasible set. When $J(x)\mathbf{n}(x) = 0$ for all $x \in \partial K$, we have $\mathbf{n}^J(x) = \mathbf{n}(x)$ on ∂K . Consequently, the skew-projection correction reduces to the standard normal one (and \mathcal{P}_K^J coincides with the usual projection/reflection correction).

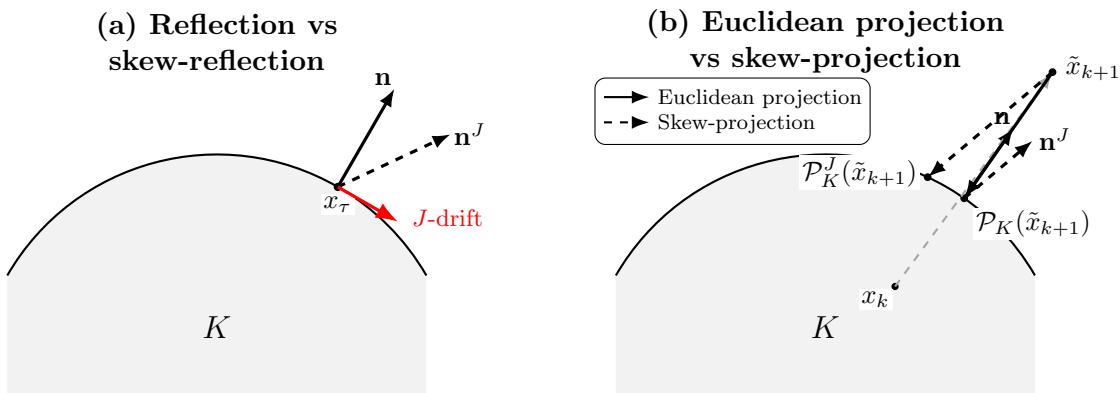


Figure 1: Geometric interpretation of the dynamics. **(a)** At a boundary point $x_\tau \in \partial K$, standard reflection pushes along the normal direction \mathbf{n} , whereas skew-reflection pushes along \mathbf{n}^J ; when $J(x)\mathbf{n}(x) = 0$ on ∂K , $\mathbf{n}^J = \mathbf{n}$ on ∂K . Otherwise, \mathbf{n}^J generally introduces an additional tangential component (red) induced by the skew field. **(b)** Starting from x_k , the algorithm proposes an unconstrained step to \tilde{x}_{k+1} and then maps it back to K via either the Euclidean projection \mathcal{P}_K (solid) or an oblique correction in the direction $-\mathbf{n}^J$ leading to the skew-projection \mathcal{P}_K^J (dashed).

Although non-asymptotic convergence analysis for SRNLD and SRNLMC are obtained in Du et al. (2025) and it is shown that by adding the skew-symmetric matrix J , acceleration can be achieved in the context of constrained sampling, the analysis in Du et al. (2025) does not shed any light on *how to design and construct* the skew-symmetric matrix J to achieve good performance in practice. Indeed, even in the unconstrained setting, how to choose the skew-symmetric matrix is a challenging problem with very limited understanding. In the unconstrained setting, when the objective f is quadratic, X_t becomes a Gaussian process. Using the rate of convergence of the covariance of X_t as the criterion, Hwang et al. (1993) showed that $J \equiv 0$ is the worst choice. Lelièvre et al. (2013) proved the existence of the optimal skew-symmetric matrix J to accelerate the convergence to equilibrium via maximizing spectral gaps. Wu et al. (2014) proposed two approaches to designing J for optimal convergence of Gaussian diffusion and they also compared their algorithms with the one

in Lelievre et al. (2013). See also Guillin and Monmarché (2016) for related results. For non-quadratic objectives, the literature is sparse but reveals consistent acceleration benefits. Gao et al. (2020) demonstrated reduced recurrence times in non-convex optimization via non-reversible Langevin dynamics, by investigating the metastability behavior. Le Peutrec and Michel (2020) derived sharp spectral asymptotics for non-reversible metastable diffusions, and confirmed faster transitions via Eyring-Kramers formulas. Furthermore, Lee and Seo (2022) proved accelerated Eyring-Kramers formulas for non-reversible Gibbs-sampled diffusions using novel capacity estimation. Collectively, these works validate that breaking reversibility enhances convergence, yet the optimal J for general non-quadratic objectives remains an open challenge. To the best of our knowledge, we are the first to study how to choose and design the skew-symmetric matrix J in the context of constrained sampling. We adopt a different approach than Du et al. (2025) to study SRNLD through the lens of *large deviations theory* that can help us better understand how to design and construct the skew-symmetric matrix J effectively.

In this paper, we focus on the study of large deviations of the *empirical measure* for SRNLD $(X_t)_{t \geq 0}$ in (1.5):

$$L_t := \frac{1}{t} \int_0^t \delta_{X_s} ds, \quad (1.9)$$

such that $L_t \in \mathcal{P}(K)$, where $\mathcal{P}(K)$ denotes the space of probability measures on K . Since (1.5) admits μ as a unique invariant distribution, by ergodic theorem, $L_t \rightarrow \mu$ as $t \rightarrow \infty$ a.s. Informally speaking, the large deviations theory concerns the small probability of the rare events $\mathbb{P}(L_t \simeq \nu)$ where $\nu \neq \mu$. Typically, this probability is exponentially small, i.e. $\mathbb{P}(L_t \simeq \nu) = e^{-tI(\nu)+o(t)}$ as $t \rightarrow \infty$, where $I(\nu) \geq 0$ is known as the *rate function* with $I(\nu) = 0$ if and only if $\nu = \mu$. More formally speaking, the empirical measures $(L_t)_{t \geq 0}$ satisfy a *large deviation principle* (LDP) in $\mathcal{P}(K)$ (equipped with the weak topology) with speed t and rate function $I : \mathcal{P}(K) \rightarrow [0, \infty]$ if I is non-negative, lower semicontinuous and for any measurable set A , we have

$$-\inf_{\nu \in A^\circ} I(\nu) \leq \liminf_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}(L_t \in A) \leq \limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}(L_t \in A) \leq -\inf_{\nu \in \bar{A}} I(\nu), \quad (1.10)$$

where A° denotes the interior of A and \bar{A} its closure, and the rate function $I(\nu)$ is said to be good if it has compact level sets. We refer to Varadhan (1984); Dembo and Zeitouni (1998) for a more general definition of large deviation principle, as well as its theory and applications.

Large deviations theory can help us better understand the convergence to the target distribution. When we have variants of Langevin dynamics that target the same Gibbs distribution μ , they might lead to different rate functions, and if we can show the rate function for SRNLD X_t given in (1.5) is greater than that for the RLD (1.4), then as $t \rightarrow \infty$, SRNLD is more concentrated around the target distribution μ , which indicates acceleration.

The Donsker-Varadhan theory (Donsker and Varadhan, 1975a,b, 1976, 1983) establishes a large deviation principle with a rate function $I(\nu)$ for the empirical measure L_t when the underlying process X_t is a Markov process under suitable conditions. However, the Donsker-Varadhan theory does not apply to Langevin SDEs directly because overdamped Langevin SDEs and its variants on \mathbb{R}^d live on a non-compact set that requires redesign and rethinking

(Ferré and Stoltz, 2020; Yao et al., 2025). Since reflected Langevin dynamics is reversible and on a compact domain, its large deviation principle follows from the literature. This leads to the natural question whether one can establish large deviations for skew-reflected non-reversible Langevin dynamics. It turns out that it is a much harder problem for the following technical reasons:

- (i) **General Boundary Conditions:** For an arbitrary skew-symmetric matrix field $J(x)$, the domain $D(\mathcal{L}_J)$ of the infinitesimal generator $\mathcal{L}_J u(x) = \Delta u(x) - \langle (I + J(x))\nabla f(x), \nabla u(x) \rangle$ consists of functions $u \in C^2(K)$ satisfying the boundary condition $\langle \nabla u(x), (I + J(x))\mathbf{n}(x) \rangle = 0$ on ∂K . This is a generalized oblique derivative condition, which can be significantly more complex to analyze than standard boundary conditions such as Neumann or Dirichlet conditions. In contrast, the classical framework of Donsker and Varadhan (1975a) applies only to the case $J \equiv 0$, where the Markov process is reversible, whereas our setting with a non-zero $J(x)$ introduces a non-reversible reflected process that falls outside their framework.
- (ii) **Operator Decomposition and Adjoints:** The standard approach to decompose \mathcal{L}_J into a symmetric part $\mathcal{L}_S = \frac{1}{2}(\mathcal{L}_J + \mathcal{L}_J^*)$ and a skew-symmetric part $\mathcal{L}_A = \frac{1}{2}(\mathcal{L}_J - \mathcal{L}_J^*)$ (where \mathcal{L}_J^* is the adjoint in $L^2(K, d\mu)$ with respect to the invariant measure μ) requires a well-defined adjoint \mathcal{L}_J^* . The calculation of \mathcal{L}_J^* via integration by parts involves boundary terms. The nature of these boundary terms, and consequently the precise form and domain of \mathcal{L}_J^* , are dictated by the original oblique boundary condition.
- (iii) **Properties of Decomposed Operators:** For the LDP rate function decomposition methods (e.g., as in Ferré and Stoltz (2020)) to be applicable, it is highly desirable that \mathcal{L}_S is self-adjoint and \mathcal{L}_A is anti-self-adjoint, ideally on the same domain $D(\mathcal{L}_J)$ or at least on a common core. If the boundary conditions for $D(\mathcal{L}_J)$ and $D(\mathcal{L}_J^*)$ (which arises from the oblique condition) do not align simply, establishing these self-adjointness and anti-self-adjointness properties for \mathcal{L}_S and \mathcal{L}_A becomes a non-trivial task. The resulting \mathcal{L}_S and \mathcal{L}_A might not have domains that are easy to work with for subsequent spectral analysis or variational arguments.

In this paper, we will overcome these technical challenges. The key insight is the condition $J(x)\mathbf{n}(x) = 0$ on ∂K we impose (Assumption 2 in this work) which is crucial because it simplifies the oblique boundary condition $\langle \nabla u, (I + J)\mathbf{n} \rangle = 0$ to the standard Neumann condition $\langle \nabla u, \mathbf{n} \rangle = 0$. For operators with Neumann boundary conditions, the theory of adjoints, self-adjoint extensions, and spectral properties is much better developed and more tractable. This simplification ensures that \mathcal{L}_J^* can be cleanly characterized, leading to well-behaved \mathcal{L}_S and \mathcal{L}_A that are indeed self-adjoint and anti-self-adjoint respectively on $D(\mathcal{L}_J)$. This, in turn, allows for a more direct application and adaptation of the LDP framework and rate function decomposition techniques found in the literature such as Ferré and Stoltz (2020). Without this simplification, one might need to delve into more advanced partial differential equation (PDE) theory for oblique boundary value problems or find alternative approaches for the LDP analysis. As a by-product, this simplification also provides us a natural choice for the skew-symmetric matrix, which will be discussed for the examples including the ball constraint, and more generally the constrained domains characterized by

sublevel sets that include special cases such as the smoothed ℓ_p ball constraint in Section 2 which turns out to work well empirically (Section 3).

With this insight, in Section 2, we will establish a large deviation principle for the empirical measures of SRNLD for a class of skew-symmetric matrices J by adopting the Gärtner-Ellis framework (Theorem 7). Moreover, we show that the acceleration can be achieved by comparing the rate function in the presence of the skew-symmetric matrix J and that of RLD without the skew-symmetric matrix (Proposition 11). In addition, we rigorously prove that these non-reversible dynamics lead to a reduction in the asymptotic variance for time-average estimators (Theorem 15). The class of skew-symmetric matrices J can be constructed depending on the compact domain K (Assumptions 2 and 3). As a by-product, it provides us a guidance how to design and construct the skew-symmetric matrix J to achieve better performance in the context of constrained sampling, which bridges a gap between theory and practice (Section 3). In particular, we conduct numerical experiments including a toy example on truncated multivariate normal distribution, constrained Bayesian linear regression and constrained Bayesian logistic regression for both the ball constraint and the smoothed ℓ_p constraint using both synthetic data and real data. Our numerical experiments demonstrate the efficiency of the algorithmic design, and show superior performance. Finally, we conclude in Section 4. We provide a summary of notations in Appendix A. The proofs of technical lemmas will be provided in Appendix B and additional technical derivations will be presented in Appendix C.

2. Main Results

We consider a *skew-reflected non-reversible Langevin dynamics* (SRNLD) on a compact, connected domain $K \subset \mathbb{R}^d$ with boundary $\partial K \in C^{2,\alpha}$ for some $\alpha \in (0, 1)$:

$$dX_t = -(I + J(X_t))\nabla f(X_t)dt + \sqrt{2}dW_t, \quad X_t \in K, \quad (2.1)$$

subject to reflecting boundary conditions. The infinitesimal generator is

$$\mathcal{L}_J u(x) = \Delta u(x) - \langle (I + J(x))\nabla f(x), \nabla u(x) \rangle, \quad (2.2)$$

with the domain

$$D(\mathcal{L}_J) = \{ u \in C^{2,\alpha}(K) : \langle \nabla u(x), (I + J(x))\mathbf{n}(x) \rangle = 0 \text{ on } \partial K \}.$$

The empirical measure is $L_t = \frac{1}{t} \int_0^t \delta_{X_s} ds \in \mathcal{P}(K)$, the space of probability measures on K . We seek an LDP for $(L_t)_{t \geq 0}$ in the weak topology on $\mathcal{P}(K)$. We introduce the following assumptions that will be used throughout the rest of the paper.

Assumption 1 (Regularity) *The domain $K \subset \mathbb{R}^d$ is compact and connected, and its boundary satisfies $\partial K \in C^{2,\alpha}$ for some $\alpha \in (0, 1)$. The potential satisfies $f \in C^{2,\alpha}(K)$. The matrix field satisfies $J \in C^{1,\alpha}(K) \cap L^\infty(K)$ and $J^\top = -J$ on K .*

Assumption 2 (Simplified Boundary Interaction) *On the boundary ∂K , $J(x)\mathbf{n}(x) = 0$, where $\mathbf{n}(x)$ is the outward unit normal vector.*

Degeneracy under Assumption 2 Under Assumption 2, we have $J(x)\mathbf{n}(x) = 0$ for all $x \in \partial K$. Consequently, on ∂K the modified reflection direction $\mathbf{n}^J(x)$ coincides with the outward normal $\mathbf{n}(x)$, so the skew-reflection reduces to the standard (normal) reflection. We keep the notation \mathbf{n}^J for consistency with the general SRNLD/SRNLMC formulation. In particular, $\langle \mathbf{n}^J(x), \mathbf{n}(x) \rangle = 1$ on ∂K .

Assumption 3 *In the interior K° , the matrix field is divergence-free in the sense that $\nabla \cdot J = 0$ on K° .*

The assumptions introduced above serve distinct roles in establishing the theoretical framework:

- **Invariance of μ (Lemma 1):** The result that $\mu(dx) \propto e^{-f(x)}dx$ is the invariant measure relies primarily on the skew-symmetry of J (Assumption 1) and the divergence-free condition $\nabla \cdot J = 0$ (Assumption 3). These conditions ensure that the additional drift term vanishes in the interior.
- **Existence of LDP (Theorem 7):** The establishment of the LDP via the Gärtner-Ellis theorem relies heavily on the compactness of K and the regularity of the coefficients (Assumption 1). Furthermore, Assumption 2 simplifies the domain of the generator to standard Neumann boundary conditions, which facilitates the use of standard spectral theory for elliptic operators on compact domains to prove the existence and differentiability of the scaled cumulant generating function (SCGF) $\lambda(g)$.
- **Rate Function Decomposition and Variance Reduction (Proposition 11 & Theorem 15):** The algebraic decomposition of the rate function and the variance reduction result depend on the specific structure of the adjoint operator \mathcal{L}_J^* . This calculation requires the full bundle of assumptions: Assumption 3 ($\nabla \cdot J = 0$) eliminates interior terms in the integration by parts, while Assumption 2 ($J\mathbf{n} = 0$) eliminates the boundary terms, ensuring that \mathcal{L}_S is self-adjoint and \mathcal{L}_A is anti-self-adjoint on the domain $D(\mathcal{L}_J)$.

Next, we discuss how to construct $J(x)$ to satisfy these assumptions and their implications.

2.1 Construction of $J(x)$ and Implications of Assumptions

The skew-symmetric matrix field $J(x)$ is chosen to satisfy three key assumptions:

- (i) **Regularity (Assumption 1):** $J \in C^{1,\alpha}(K) \cap L^\infty(K)$ and $J(x) = -(J(x))^\top$.
- (ii) **Simplified Boundary Interaction (Assumption 2):** $J(x)\mathbf{n}(x) = 0$ on ∂K .
- (iii) **Divergence-Free Condition (Assumption 3):** $\nabla \cdot J(x) = 0$ for $x \in K^\circ$.

Constructing $J(x)$ to satisfy all three conditions simultaneously is non-trivial.

2.1.1 IN 3D SPACE

A common way to ensure $J(x)$ is skew-symmetric (fulfilling part of Assumption 1) is to define its action via an axial vector field $\mathbf{k}(x)$: $J(x)\mathbf{w} = \mathbf{k}(x) \times \mathbf{w}$.

To satisfy Assumption 3 ($\nabla \cdot J = 0$), we require the axial vector field $\mathbf{k}(x)$ to be irrotational (curl-free) in K° , i.e., $\nabla \times \mathbf{k}(x) = \mathbf{0}$. This implies $\mathbf{k}(x)$ can be expressed as the gradient of a scalar potential $\phi(x)$, i.e., $\mathbf{k}(x) = \nabla\phi(x)$. To satisfy Assumption 2 ($J(x)\mathbf{n}(x) = 0$) on ∂K , we require $\mathbf{k}(x) \times \mathbf{n}(x) = \mathbf{0}$ on ∂K , which means $\mathbf{k}(x)$ must be parallel to $\mathbf{n}(x)$ on the boundary. Therefore, in three dimensions, it suffices to find a scalar potential ϕ such that

$$\mathbf{k}(x) := \nabla\phi(x)$$

has the required Hölder regularity (so that the resulting matrix field $J(x)$ satisfies Assumption 1). Moreover, on the boundary ∂K we require $\nabla\phi(x)$ to be parallel to the outward normal $\mathbf{n}(x)$, i.e.

$$\nabla\phi(x) \times \mathbf{n}(x) = 0, \quad x \in \partial K.$$

For instance, choosing $\mathbf{k}(x) = sx$ (with $s \in \mathbb{R}$) yields $\nabla \times \mathbf{k}(x) = \mathbf{0}$. If K is a ball centered at the origin, then $\mathbf{n}(x)$ is parallel to x on ∂K , hence $\mathbf{k}(x) \parallel \mathbf{n}(x)$ holds. Consequently, the corresponding $J(x)$ (analogous to $J_s(x)$ in Section 3) satisfies the required assumptions.

2.1.2 IN A HIGHER-DIMENSIONAL BALL ($d > 3$)

For our high-dimensional numerical experiments in Section 3, we additionally employ a simple block-diagonal construction on a ball, which yields a more structured nonreversible drift and is easy to implement.

Let K be a ball in \mathbb{R}^d centered at the origin, which corresponds to the ℓ_2 -norm constraints, that are frequently encountered in machine learning applications, such as Bayesian ridge regression (Brosse et al., 2017; Gürbüzbalaban et al., 2024a). The outward unit normal vector on the boundary ∂K is $\mathbf{n}(x) = x/\|x\|_2$. The boundary condition in Assumption 2, $J(x)\mathbf{n}(x) = 0$, thus simplifies to $J(x)x = 0$ for $x \in \partial K$. We can construct a $J(x)$ that meets all requirements using a block-diagonal structure.

Specifically, when the dimension d is a multiple of 3, say $d = 3m$, we construct $J(x)$ in the following matrix form:

$$J(x) = \begin{pmatrix} J^{(1)}(x_{1:3}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J^{(2)}(x_{4:6}) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J^{(m)}(x_{3m-2:3m}) \end{pmatrix},$$

where $x_{3i-2:3i} = (x_{3i-2}, x_{3i-1}, x_{3i})^\top$ and $\mathbf{0}$ represents a 3×3 zero matrix. Each 3D block $J^{(i)}$ is constructed analogously to the 3D case. Specifically, the local axial vector $\mathbf{k}^{(i)}$ plays the role of the vector field \mathbf{k} restricted to the coordinates of the i -th block. For example, we may choose $\mathbf{k}^{(i)}(x) = s_i \cdot x_{3i-2:3i}$, where s_i is a scalar constant. The action of such a block is $J^{(i)}\mathbf{w} = (s_i \cdot x_{3i-2:3i}) \times \mathbf{w}$.

This construction rigorously satisfies all three assumptions for a spherical domain:

1. **Skew-symmetry:** Since each block $J^{(i)}$ is skew-symmetric, the full block-diagonal matrix $J(x)$ is also skew-symmetric.

2. **Boundary condition** ($J(x)x = 0$): For any vector $x = (x_{1:3}^\top, \dots, x_{3m-2:3m}^\top)^\top$, the action $J(x)x$ results in a vector where each 3-dimensional component is $J^{(i)}x_{3i-2:3i} = (s_i x_{3i-2:3i}) \times x_{3i-2:3i} = \mathbf{0}$. Thus, $J(x)x = \mathbf{0}$ for all x , which satisfies the boundary condition.
3. **Divergence-free** ($\nabla \cdot J = 0$): The divergence of each 3D block with respect to its own coordinates is zero. Due to the block-diagonal structure with no cross-dependencies between blocks, the divergence of the full matrix $J(x)$ is also zero.

This block-diagonal construction is what we will employ in our high-dimensional numerical experiments (see Section 3.3 with $d = 9$). Finally, if the dimension d is not a multiple of 3, one can combine 3D blocks with zero blocks for the remaining dimensions, which still satisfies all the assumptions.

2.1.3 CONSTRAINED DOMAIN CHARACTERIZED BY SUBLEVEL SETS IN \mathbb{R}^d

We now present a constructive procedure to design the skew-symmetric matrix field $J(x)$ for a general class of constrained domains defined by sublevel sets. Constraints characterized by sublevel sets (or equivalently, functional inequalities) are ubiquitous in the sampling and optimization literature, encompassing standard convex bodies, polyhedra, and ℓ_p -balls (Schmidt, 2005; Kook et al., 2022; Gürbüzbalaban et al., 2024a). This generalizes the explicit construction provided for the ℓ_2 -ball example.

Let $g \in C^{2,\alpha}(\mathbb{R}^d)$ be a confining potential. Fix a level $\lambda > \inf_{x \in \mathbb{R}^d} g(x)$ and let

$$K := \{x \in \mathbb{R}^d : g(x) \leq \lambda\}.$$

We assume that K is bounded and that $\nabla g(x) \neq 0$ for all $x \in \partial K = \{g = \lambda\}$. Under these conditions, ∂K is a $C^{2,\alpha}$ hypersurface, and the outward unit normal vector is given by $\mathbf{n}(x) = \nabla g(x) / \|\nabla g(x)\|$.

The key to satisfying the boundary condition $J(x)\mathbf{n}(x) = 0$ is to construct an auxiliary potential ψ whose gradient aligns with the normal vector on the boundary. Choose an arbitrary function $h \in C^{2,\alpha}(\mathbb{R}^d)$ such that $h(x) \neq 0$ on ∂K (e.g., $h(x) \equiv 1$ or $h(x) = 1 + \|x\|^2$). Define the potential $\psi : K \rightarrow \mathbb{R}$ by:

$$\psi(x) := (\lambda - g(x)) h(x). \tag{2.3}$$

On the boundary ∂K , where $g(x) = \lambda$, the gradient of ψ satisfies:

$$\nabla \psi(x) = -\nabla g(x) h(x) + (\lambda - g(x)) \nabla h(x) = -h(x) \nabla g(x).$$

Crucially, this implies that $\nabla \psi(x)$ is parallel to the normal vector $\mathbf{n}(x)$ everywhere on ∂K .

We define $J(x)$ by decomposing the d -dimensional space into independent 3-dimensional subspaces (with potentially one or two dimensions left over if d is not a multiple of 3). Let $m = \lfloor d/3 \rfloor$. For each $\ell = 1, \dots, m$, consider the coordinate triplet $(a_\ell, b_\ell, c_\ell) = (3\ell - 2, 3\ell - 1, 3\ell)$. Define the block-specific axial vector $\mathbf{k}^{(\ell)}(x) \in \mathbb{R}^3$ by the partial derivatives of ψ :

$$\mathbf{k}^{(\ell)}(x) := \begin{pmatrix} \partial_{a_\ell} \psi(x) \\ \partial_{b_\ell} \psi(x) \\ \partial_{c_\ell} \psi(x) \end{pmatrix}.$$

Construct the $d \times d$ matrix $J^{(\ell)}(x)$ such that its only non-zero entries form a 3×3 skew-symmetric block on the indices (a_ℓ, b_ℓ, c_ℓ) . By denoting the components of the local axial vector as $\mathbf{k}^{(\ell)}(x) = (k_1^{(\ell)}, k_2^{(\ell)}, k_3^{(\ell)})^\top$, this block is given by:

$$\left(J^{(\ell)} \right)_{\{a_\ell, b_\ell, c_\ell\} \times \{a_\ell, b_\ell, c_\ell\}} = \begin{pmatrix} 0 & -k_3^{(\ell)} & k_2^{(\ell)} \\ k_3^{(\ell)} & 0 & -k_1^{(\ell)} \\ -k_2^{(\ell)} & k_1^{(\ell)} & 0 \end{pmatrix}.$$

Finally, set the full matrix field as the sum:

$$J(x) := \sum_{\ell=1}^m J^{(\ell)}(x). \quad (2.4)$$

This construction satisfies all required assumptions:

1. **Regularity:** Since $g, h \in C^{2,\alpha}$, $\psi \in C^{2,\alpha}$ and $J \in C^{1,\alpha}$. J is skew-symmetric by construction.
2. **Boundary Condition:** For any $x \in \partial K$, we showed $\nabla\psi(x) \parallel \mathbf{n}(x)$. This implies that for each block ℓ , the local vector $\mathbf{k}^{(\ell)}$ is parallel to the projection of \mathbf{n} onto those coordinates. The cross-product structure ensures that $J^{(\ell)}(x)\mathbf{n}(x) = 0$. Summing these gives $J(x)\mathbf{n}(x) = 0$.
3. **Divergence-Free:** The divergence of a curl field is identically zero. Since each block is constructed as the curl of a gradient field ($\nabla\psi$), we have $\nabla \cdot J = 0$ in K° .

2.1.4 SMOOTHED ℓ_p BALL CONSTRAINT

Next, we apply the discussions in the previous section to the example of a smoothed ℓ_p ball sublevel-set constraint in \mathbb{R}^d . Constraints of this type are widely encountered in high-dimensional statistics and machine learning for regularization, such as in Bayesian Lasso ($p = 1$) and ridge ($p = 2$) regression (Schmidt, 2005; Gürbüzbalaban et al., 2024a). Fix $d \geq 3$, $p \geq 1$ and $\varepsilon > 0$, and define the smooth convex potential

$$g(x) := \sum_{i=1}^d (x_i^2 + \varepsilon^2)^{\frac{p}{2}}, \quad x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d. \quad (2.5)$$

Then $\min_{x \in \mathbb{R}^d} g(x) = d\varepsilon^p$ (attained at $x = 0$). For any $\lambda > d\varepsilon^p$, we consider the sublevel-set domain

$$K := \{x \in \mathbb{R}^d : g(x) \leq \lambda\}. \quad (2.6)$$

We have $\partial K = \{g = \lambda\}$ and $\nabla g(x) \neq 0$ on ∂K since $\nabla g(x) = 0$ if and only if $x = 0$, while $0 \in K^\circ$ because $g(0) = d\varepsilon^p < \lambda$. The outward unit normal vector on ∂K is

$$\mathbf{n}(x) = \frac{\nabla g(x)}{\|\nabla g(x)\|}, \quad x \in \partial K, \quad (2.7)$$

where, for $i = 1, \dots, d$,

$$\partial_i g(x) = p x_i (x_i^2 + \varepsilon^2)^{\frac{p}{2}-1}. \quad (2.8)$$

Define

$$\psi(x) := \lambda - g(x). \quad (2.9)$$

For each triple of indices (a, b, c) with $1 \leq a < b < c \leq d$, define the associated 3-vector

$$\mathbf{k}^{(a,b,c)}(x) := \begin{pmatrix} \partial_a \psi(x) \\ \partial_b \psi(x) \\ \partial_c \psi(x) \end{pmatrix}. \quad (2.10)$$

Equivalently, for any $r \in \{a, b, c\}$,

$$\partial_r \psi(x) = -\partial_r g(x). \quad (2.11)$$

For the choice (2.5), we have for each $r \in \{1, \dots, d\}$,

$$\partial_r \psi(x) = -\partial_r g(x) = -p x_r (x_r^2 + \varepsilon^2)^{\frac{p}{2}-1}. \quad (2.12)$$

Hence, for any triple (a, b, c) , we have:

$$\mathbf{k}^{(a,b,c)}(x) := \begin{pmatrix} k_1^{(a,b,c)}(x) \\ k_2^{(a,b,c)}(x) \\ k_3^{(a,b,c)}(x) \end{pmatrix} = -p \begin{pmatrix} x_a (x_a^2 + \varepsilon^2)^{\frac{p}{2}-1} \\ x_b (x_b^2 + \varepsilon^2)^{\frac{p}{2}-1} \\ x_c (x_c^2 + \varepsilon^2)^{\frac{p}{2}-1} \end{pmatrix}. \quad (2.13)$$

In particular, the components satisfy $k_j^{(a,b,c)}(x) = \partial_{\iota_j} \psi(x)$, where $(\iota_1, \iota_2, \iota_3) = (a, b, c)$.

We now select a family of disjoint coordinate triples that cover as many coordinates as possible. Let

$$m := \left\lfloor \frac{d}{3} \right\rfloor, \quad (a_\ell, b_\ell, c_\ell) := (3\ell - 2, 3\ell - 1, 3\ell), \quad \ell = 1, \dots, m.$$

(If $d \not\equiv 0 \pmod{3}$, the remaining 1 or 2 coordinates will be left without a swirl block.)

For each $\ell = 1, \dots, m$, define a $d \times d$ matrix field $J^{(\ell)}(x)$ supported on the coordinates (a_ℓ, b_ℓ, c_ℓ) by inserting the 3×3 cross-product matrix generated by $\mathbf{k}^{(a_\ell, b_\ell, c_\ell)}(x)$ into those rows and columns:

$$\left[J^{(\ell)}(x) \right]_{\{a_\ell, b_\ell, c_\ell\} \times \{a_\ell, b_\ell, c_\ell\}} = \begin{pmatrix} 0 & -k_3^{(a_\ell, b_\ell, c_\ell)}(x) & k_2^{(a_\ell, b_\ell, c_\ell)}(x) \\ k_3^{(a_\ell, b_\ell, c_\ell)}(x) & 0 & -k_1^{(a_\ell, b_\ell, c_\ell)}(x) \\ -k_2^{(a_\ell, b_\ell, c_\ell)}(x) & k_1^{(a_\ell, b_\ell, c_\ell)}(x) & 0 \end{pmatrix}, \quad (2.14)$$

and set all other entries of $J^{(\ell)}(x)$ equal to 0. In particular, by (2.12)–(2.13), for the smoothed ℓ_p ball example,

$$\begin{aligned} k_1^{(a_\ell, b_\ell, c_\ell)}(x) &= -p x_{a_\ell} (x_{a_\ell}^2 + \varepsilon^2)^{\frac{p}{2}-1}, \\ k_2^{(a_\ell, b_\ell, c_\ell)}(x) &= -p x_{b_\ell} (x_{b_\ell}^2 + \varepsilon^2)^{\frac{p}{2}-1}, \\ k_3^{(a_\ell, b_\ell, c_\ell)}(x) &= -p x_{c_\ell} (x_{c_\ell}^2 + \varepsilon^2)^{\frac{p}{2}-1}. \end{aligned}$$

We define

$$J(x) := \sum_{\ell=1}^m J^{(\ell)}(x), \quad (2.15)$$

where $J^{(\ell)}(x)$ is defined in (2.14). By construction, the matrix field $J(x)$ in (2.15) satisfies the required theoretical conditions. First, for $\varepsilon > 0$, the function g is smooth, hence $\psi = \lambda - g$ is smooth as well; therefore each component $\partial_r \psi$ is C^1 on K , and so $J \in C^1(K)$ (Assumption 1).

Second, on $\partial K = \{g = \lambda\} = \{\psi = 0\}$ we have $\nabla \psi = -\nabla g$, so $\nabla \psi(x)$ is colinear with the outward unit normal $\mathbf{n}(x) = \nabla g(x)/\|\nabla g(x)\|$ (indeed anti-parallel). Since each local axial vector $\mathbf{k}^{(a_\ell, b_\ell, c_\ell)}(x)$ is defined by selecting components of $\nabla \psi(x)$, it follows that the restricted vectors $(n_{a_\ell}, n_{b_\ell}, n_{c_\ell})^\top$ and $\mathbf{k}^{(a_\ell, b_\ell, c_\ell)}(x)$ are colinear on ∂K . Consequently, on ∂K the 3×3 cross-product block generated by $\mathbf{k}^{(a_\ell, b_\ell, c_\ell)}(x)$ annihilates $\mathbf{n}(x)$ in those coordinates, and hence

$$J(x)\mathbf{n}(x) = 0, \quad x \in \partial K$$

(Assumption 2).

Finally, each block $J^{(\ell)}$ is divergence-free. Writing $(a, b, c) = (a_\ell, b_\ell, c_\ell)$ and $\mathbf{k} = (k_1, k_2, k_3)^\top = (\partial_a \psi, \partial_b \psi, \partial_c \psi)^\top$, the nonzero entries are $J_{ab} = -k_3$, $J_{ac} = k_2$, $J_{bc} = -k_1$ (and skew-symmetry). Thus, for example, since $J_{aj} = 0$ for all $j \notin \{b, c\}$,

$$\left(\nabla \cdot J^{(\ell)}\right)_a = \sum_{j=1}^d \partial_j J_{aj} = \partial_b J_{ab} + \partial_c J_{ac} = -\partial_b k_3 + \partial_c k_2 = -(\partial_b \partial_c \psi) + (\partial_c \partial_b \psi) = 0,$$

and similarly $(\nabla \cdot J^{(\ell)})_b = (\nabla \cdot J^{(\ell)})_c = 0$, while all other components vanish trivially. Therefore $\nabla \cdot J^{(\ell)} = 0$ for each ℓ , and summing over ℓ yields $\nabla \cdot J = 0$ (Assumption 3).

Standard ℓ_p -ball constraint It is worth noting the conditions under which one can set the smoothing parameter $\varepsilon = 0$ to recover the standard ℓ_p -ball constraint. In the limit $\varepsilon \rightarrow 0$, the components $k_j(x)$ become proportional to $x|x|^{p-2}$ (interpreted as $\text{sgn}(x)|x|^{p-1}$).

- If $p \geq 2$, the function $x \mapsto x|x|^{p-2}$ belongs to $C^1(\mathbb{R})$ (and $C^2(\mathbb{R})$ if $p \geq 3$). In this case, the resulting matrix field $J(x)$ satisfies the $C^{1,\alpha}$ regularity required by Assumption 1, and thus we can set $\varepsilon = 0$.
- If $1 \leq p < 2$, the derivative of $x|x|^{p-2}$ behaves like $|x|^{p-2}$, which becomes singular at $x = 0$. Therefore, for $1 \leq p < 2$, a strictly positive smoothing parameter $\varepsilon > 0$ is necessary to ensure the validity of our theoretical framework.

2.2 Preliminaries: Ergodicity and Invariant Measure

We first introduce the following lemma from Du et al. (2025) which guarantees that μ is the unique invariant distribution for $(X_t)_{t \geq 0}$ in (2.1).

Lemma 1 (Existence and Uniqueness of Invariant Measure) *Under Assumptions 1, 2 and 3, the process $(X_t)_{t \geq 0}$ defined by Eq. (2.1) with Neumann boundary conditions has a unique invariant probability measure $\mu \in \mathcal{P}(K)$. This measure has a density with respect to Lebesgue measure dx on K , given by $d\mu(x) = \frac{1}{Z} e^{-f(x)} dx$, where $Z := \int_K e^{-f(x)} dx$. The process $(X_t)_{t \geq 0}$ is ergodic with respect to μ .*

2.3 The Gärtner-Ellis Framework

We will establish LDP using the Gärtner-Ellis theorem. In the Gärtner-Ellis framework, the key object is the *scaled cumulant generating function* (SCGF).

Definition 2 (Scaled Cumulant Generating Function) For $g \in C(K)$ (the space of continuous functions on K), the SCGF is defined as:

$$\lambda(g) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g(X_s) ds \right) \right],$$

whenever the limit exists. We need to show that $\lambda(g)$ exists, is finite, convex, Gateaux-differentiable, and does not depend on the starting point $x \in K$. These properties are not merely technical: existence and finiteness ensure that $\lambda(g)$ is a well-defined long-time growth rate for the exponential moment; convexity and Gateaux differentiability are needed to identify the associated rate function via the Legendre–Fenchel transform and to characterize the unique “tilted” invariant measure through the gradient $\nabla \lambda(g)$; finally, independence of the initial condition guarantees that $\lambda(g)$ describes an intrinsic asymptotic quantity of the dynamics rather than a transient effect. This naturally leads to the study of the Feynman–Kac semigroup and its spectral properties.

2.3.1 FEYNMAN-KAC SEMIGROUP AND SPECTRAL PROPERTIES

Define the Feynman-Kac semigroup $P_t^g : C(K) \rightarrow C(K)$ by

$$(P_t^g \phi)(x) = \mathbb{E}^x \left[\phi(X_t) \exp \left(\int_0^t g(X_s) ds \right) \right].$$

The SCGF $\lambda(g)$ is related to the principal eigenvalue of the operator $\mathcal{L}_J + g$. We will prove for any $g \in C(K)$:

1. $\lambda(g)$ exists, is finite, and is independent of $x \in K$.
2. $\lambda(g)$ is the principal eigenvalue of the operator $\mathcal{L}_J + g$ (acting on functions satisfying the boundary condition). That is, there exists a unique positive eigenfunction $h_g \in C(K)$ (unique up to scaling) such that $(\mathcal{L}_J + g)h_g = \lambda(g)h_g$.
3. The function $\lambda : C(K) \rightarrow \mathbb{R}$ is convex and continuous.
4. The function $\lambda : C(K) \rightarrow \mathbb{R}$ is Gateaux-differentiable. More precisely, for any $g, \psi \in C(K)$ the limit

$$D\lambda(g)[\psi] = \lim_{\varepsilon \rightarrow 0} \frac{\lambda(g + \varepsilon\psi) - \lambda(g)}{\varepsilon}$$

exists and is given by

$$D\lambda(g)[\psi] = \int_K \psi(x) \nu_g(dx),$$

where ν_g is the probability measure built from the principal right/left eigenfunctions.

Eigenfunction representation of the derivative and the tilted measure Let us provide a more specific explanation of the 4th paragraph above. Let $h_g > 0$ and $h_g^* > 0$ be the principal right and left eigenfunctions of $\mathcal{L}_J + g$ and $(\mathcal{L}_J + g)^*$, respectively, normalized by $\int_K h_g(x) h_g^*(x) dx = 1$. Equivalently, in terms of the Feynman–Kac semigroup and its $L^2(K, dx)$ -adjoint,

$$P_t^g h_g = e^{\lambda(g)t} h_g, \quad P_t^{g,*} h_g^* = e^{\lambda(g)t} h_g^*,$$

and therefore for every $\varphi \in C(K)$,

$$\int_K (P_t^g \varphi)(x) h_g^*(x) dx = e^{\lambda(g)t} \int_K \varphi(x) h_g^*(x) dx.$$

Define

$$\nu_g(dx) := h_g(x) h_g^*(x) dx \quad (\text{so that } \nu_g(K) = 1).$$

Then $D\lambda(g)[\psi] = \int_K \psi d\nu_g$ as claimed. Moreover, ν_g is the invariant probability measure of the Doob h -transform of the Feynman–Kac semigroup:

$$\tilde{P}_t^g \varphi(x) := e^{-\lambda(g)t} \frac{1}{h_g(x)} P_t^g(h_g \varphi)(x), \quad \varphi \in C(K).$$

Indeed, using $P_t^g h_g = e^{\lambda(g)t} h_g$ and $\int_K (P_t^g \phi) h_g^* dx = e^{\lambda(g)t} \int_K \phi h_g^* dx$, one can check that

$$\int_K \tilde{P}_t^g \varphi d\nu_g = \int_K \varphi d\nu_g,$$

so that ν_g is the unique invariant measure associated with the tilted dynamics.

Lemma 3 (Existence of $\lambda(g)$ and a principal eigenfunction) *For any $g \in C(K)$, the Feynman–Kac semigroup P_t^g defined by*

$$(P_t^g \phi)(x) = \mathbb{E}^x \left[\phi(X_t) \exp \left(\int_0^t g(X_s) ds \right) \right]$$

acts on $C(K)$ and has the following properties:

1. P_t^g is a compact and strongly positive operator on $C(K)$ for each $t > 0$.
2. For each fixed $t > 0$, the spectral radius $r(P_t^g) > 0$ is a simple eigenvalue of P_t^g . There exists an eigenfunction $h_{t,g} \in C(K)$ such that $h_{t,g} > 0$ on K and

$$P_t^g h_{t,g} = r(P_t^g) h_{t,g}.$$

Writing $r(P_t^g) = e^{t\lambda_t(g)}$ defines $\lambda_t(g) \in \mathbb{R}$.

3. There exist $\lambda(g) \in \mathbb{R}$ and a strictly positive function $h_g \in C(K)$ such that

$$P_t^g h_g = e^{t\lambda(g)} h_g, \quad \text{for any } t \geq 0.$$

Moreover, $h_g \in D(\mathcal{L}_J + g)$ and

$$(\mathcal{L}_J + g)h_g = \lambda(g)h_g.$$

In particular, for every $t > 0$ the principal eigenfunction of P_t^g is unique up to normalization, so that $h_{t,g}$ coincides with h_g after rescaling and $\lambda_t(g) = \lambda(g)$.

4. Consequently, the limit in Definition 2 exists and equals $\lambda(g)$:

$$\lambda(g) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g(X_s) ds \right) \right],$$

and it is finite and independent of the starting point $x \in K$.

Proof The proof will be provided in Appendix B.1. ■

Lemma 4 *The function $\lambda : C(K) \rightarrow \mathbb{R}$ is convex.*

Proof The proof will be provided in Appendix B.2. ■

Lemma 5 *The function $\lambda : C(K) \rightarrow \mathbb{R}$ is continuous and Gateaux-differentiable.*

Proof The proof will be provided in Appendix B.3. ■

2.3.2 EXPONENTIAL TIGHTNESS

A key technical condition for establishing a large deviation principle (LDP) is *exponential tightness* of the family of probability measures. Exponential tightness ensures that the probability mass does not escape to regions of the space that are not compact, which is essential for the LDP to hold and for the rate function to have desirable properties such as being a good rate function (i.e., having compact level sets). We refer to (Dembo and Zeitouni, 1998, Section 1.2) for a detailed discussion. By definition (see, e.g., (Dembo and Zeitouni, 1998, Definition 1.2.18)), a family of probability measures $(\mathbb{Q}_t)_{t \geq 0}$ on a Polish space \mathcal{X} is exponentially tight if for every $A < \infty$, there exists a compact set $\Gamma_A \subset \mathcal{X}$ such that $\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{Q}_t(\Gamma_A^c) \leq -A$. In our context, $\mathcal{X} = \mathcal{P}(K)$, and $\mathbb{Q}_t(\cdot) = \mathbb{P}(L_t \in \cdot)$. We establish exponential tightness in the next following lemma.

Lemma 6 (Exponential Tightness) *The family of probability measures $(\mathbb{P}(L_t \in \cdot))_{t \geq 0}$ is exponentially tight in $\mathcal{P}(K)$ equipped with the weak topology.*

Proof The proof will be provided in Appendix B.4. ■

2.4 Large Deviation Principle

Now, we are ready to state the main result of the paper.

Theorem 7 (LDP for Empirical Measures) *Assume K is a compact metric space. Let $(X_t)_{t \geq 0}$ be the skew-reflected non-reversible Langevin dynamics on K in (2.1) with unique invariant probability measure $\mu \in \mathcal{P}(K)$. Suppose the scaled cumulant generating function $\lambda(g)$ defined for $g \in C(K)$ exists, is finite, convex, and Gateaux-differentiable on $C(K)$.*

Further, assume the family of empirical measures $(L_t)_{t \geq 0}$ is exponentially tight in $\mathcal{P}(K)$ equipped with the weak topology.

Then, the empirical measures $(L_t)_{t \geq 0}$ satisfy a large deviation principle in $\mathcal{P}(K)$ (equipped with the weak topology) with speed t and good rate function $I : \mathcal{P}(K) \rightarrow [0, \infty]$ given by

$$I(\nu) = \sup_{g \in C(K)} \left\{ \int_K g(y) d\nu - \lambda(g) \right\}.$$

The rate function $I(\nu)$ has the following properties:

- (a) $I(\nu)$ is convex and lower semi-continuous.
- (b) $I(\nu) \geq 0$ for all $\nu \in \mathcal{P}(K)$.
- (c) $I(\nu) = 0$ if and only if $\nu = \mu$.
- (d) If $I(\nu) < \infty$, then ν is absolutely continuous with respect to μ ($\nu \ll \mu$).

Proof The main statement of the LDP follows directly from the Gärtner-Ellis Theorem (e.g., (Dembo and Zeitouni, 1998, Theorem 2.3.6)).

1. The existence, finiteness, convexity, and Gateaux-differentiability of $\lambda(g)$ on $C(K)$ have been guaranteed in Lemmas 3, 4, 5.
2. Exponential tightness of $(L_t)_{t \geq 0}$ in $\mathcal{P}(K)$ is proved in Lemma 6.
3. The Gärtner-Ellis theorem then states that an LDP holds for $(L_t)_{t \geq 0}$ with speed t and rate function $I(\nu) = \lambda^*(\nu)$, where λ^* is the Legendre-Fenchel transform of $\lambda(g)$, as given in the theorem statement.
4. **Property (a):** The rate function $I(\nu)$, being the supremum of affine (and thus convex and lower semi-continuous) functions, is itself convex and lower semi-continuous. Since K is compact, $\mathcal{P}(K)$ is also compact in the weak topology. For a lower semi-continuous function on a compact space, all its level sets $\{\nu \in \mathcal{P}(K) : I(\nu) \leq \alpha\}$ are closed, and thus compact. Hence, $I(\nu)$ is a good rate function.
5. **Property (b):** $I(\nu) \geq 0$: By definition, $I(\nu) = \sup_{g \in C(K)} \{\int_K g d\nu - \lambda(g)\}$. Choose $g \equiv 0 \in C(K)$. Then the SCGF for $g \equiv 0$ is $\lambda(0) = 0$. Therefore, $I(\nu) \geq 0$.
6. **Property (c):** $I(\nu) = 0 \iff \nu = \mu$: We know $I(\mu) \geq 0$. For any $g \in C(K)$, by Jensen's inequality and the stationarity of μ :

$$\lambda(g) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g(X_s) ds \right) \right] \geq \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}^x \left[\int_0^t g(X_s) ds \right] = \int_K g d\mu,$$

where, by ergodicity, the limit is independent of x and equals expectation w.r.t. μ . This implies $\int_K g d\mu - \lambda(g) \leq 0$ for all $g \in C(K)$. Therefore,

$$I(\mu) = \sup_{g \in C(K)} \left\{ \int_K g d\mu - \lambda(g) \right\} \leq 0.$$

Combined with $I(\mu) \geq 0$, we have $I(\mu) = 0$.

For the reverse implication, we use the variational characterization of the rate function established in (Donsker and Varadhan, 1975b, Section 2). Specifically, if $I(\nu) = 0$, then by Proposition 9, we must have $\int_K \frac{\mathcal{L}_J u}{u} d\nu \geq 0$ for all positive test functions u . Considering perturbations of the form $u = 1 + \epsilon\phi$ for small ϵ and arbitrary smooth ϕ , this condition implies $\int_K \mathcal{L}_J \phi d\nu = 0$, which is precisely the definition of an invariant measure. Since μ is the *unique* invariant measure (Lemma 1), it follows that $\nu = \mu$.

7. **Property (d):** $I(\nu) < \infty \implies \nu \ll \mu$: This property is established using the spectral characterization of $\lambda(g)$ combined with elliptic regularity estimates. We proceed by contradiction.

Suppose ν is not absolutely continuous with respect to μ . Then there exists a Borel set $A \subset K$ such that $\mu(A) = 0$ but $\nu(A) > 0$. By the regularity of measures on the compact metric space K , there exists a closed set $F \subset A$ such that $\nu(F) =: \delta > 0$. Note that $\mu(F) = 0$.

Fix an arbitrary constant $M > 0$. We construct a sequence of continuous approximations. Let $d(x, F)$ be the distance from x to F . For $k \in \mathbb{N}$, define $g_k \in C(K)$ by:

$$g_k(x) = M \cdot \max(0, 1 - k \cdot d(x, F)).$$

Then $0 \leq g_k \leq M$ on K , and $g_k(x) \rightarrow M \mathbf{1}_F(x)$ pointwise.

We now control the principal eigenvalue $\lambda(g_k)$. First, by the variational representation and $0 \leq g_k \leq M$, we have $0 \leq \lambda(g_k) \leq \|g_k\|_\infty \leq M$ for all k .

Let $h_k \in C^2(K)$ be the strictly positive principal eigenfunction of $\mathcal{L}_J + g_k$ subject to the (Neumann) boundary condition, normalized by $\int_K h_k d\mu = 1$. Since μ is invariant, $\int_K \mathcal{L}_J \phi d\mu = 0$ for any sufficiently regular ϕ satisfying the boundary condition. Applying this with $\phi = h_k$ in $(\mathcal{L}_J + g_k)h_k = \lambda(g_k)h_k$ yields

$$\lambda(g_k) = \int_K g_k h_k d\mu.$$

It remains to show $\lambda(g_k) \rightarrow 0$. For this we derive a uniform L^∞ bound for h_k . Rewrite the eigenvalue equation as

$$\mathcal{L}_J h_k + c_k(x)h_k = 0, \quad c_k(x) := g_k(x) - \lambda(g_k),$$

with the Neumann boundary condition. Since $0 \leq g_k \leq M$ and $0 \leq \lambda(g_k) \leq M$, we have the uniform bound $\|c_k\|_\infty \leq 2M$.

Local Harnack inequality (interior and boundary). If $B_{2r}(x) \subset K^\circ$, then by the interior Harnack inequality (Gilbarg and Trudinger, 2001, Theorem 8.20) applied to $h_k \geq 0$ applied to $h_k \geq 0$ solving a uniformly elliptic equation with bounded (in fact C^α) coefficients and zero-order term bounded by $2M$, there exists $C_{\text{int}} > 1$ (independent of k) such that

$$\sup_{y \in B_r(x)} h_k(y) \leq C_{\text{int}} \inf_{y \in B_r(x)} h_k(y).$$

If $x \in \partial K$, then the Neumann condition is a special case of the oblique derivative boundary condition (with obliqueness bounded below), and the boundary Harnack inequality for nonnegative solutions yields constants $r_{\text{bd}} > 0$ and $C_{\text{bd}} > 1$ (depending only on the ellipticity constants (here $a_{ij} = \delta_{ij}$), the drift bound $\|b\|_{L^\infty(K)}$ with $b = (I + J)\nabla f$, the zeroth-order bound $\|c_k\|_\infty \leq 2M$, and the $C^{2,\alpha}$ -geometry of K , but independent of k) such that for all $0 < r \leq r_{\text{bd}}$,

$$\sup_{y \in B_r(x) \cap K} h_k(y) \leq C_{\text{bd}} \inf_{y \in B_r(x) \cap K} h_k(y).$$

(See (Lieberman, 1987, Corollary 3.5).)

Set $r_0 := \min\{r_{\text{bd}}, 1\}$ and $C_0 := \max\{C_{\text{int}}, C_{\text{bd}}\}$. Then for every $x \in K$ and $0 < r \leq r_0$, we have the uniform (in k) estimate

$$\sup_{y \in B_r(x) \cap K} h_k(y) \leq C_0 \inf_{y \in B_r(x) \cap K} h_k(y). \quad (2.16)$$

Global Harnack inequality via a finite chain. Fix $x^+, x^- \in K$ such that $h_k(x^+) = \sup_{x \in K} h_k(x)$ and $h_k(x^-) = \inf_{x \in K} h_k(x)$. Since K is compact and connected, there exists a finite sequence $x_0 = x^-, x_1, \dots, x_N = x^+$ such that

$$(B_{r_0/2}(x_i) \cap K) \cap (B_{r_0/2}(x_{i+1}) \cap K) \neq \emptyset, \quad i = 0, \dots, N-1,$$

where N depends only on K and r_0 . Applying (2.16) successively along the chain yields $h_k(x_{i+1}) \leq C_0 h_k(x_i)$. Hence,

$$\sup_{x \in K} h_k(x) = h_k(x^+) \leq C_0^N h_k(x^-) = C_0^N \inf_{x \in K} h_k(x).$$

Denoting $C_H := C_0^N$, we obtain the global estimate $\sup_{x \in K} h_k(x) \leq C_H \inf_{x \in K} h_k(x)$ with C_H independent of k .

Using $\int_K h_k d\mu = 1$ and $\mu(K) = 1$, we have

$$1 = \int_K h_k d\mu \geq \inf_{x \in K} h_k(x),$$

and thus $\|h_k\|_\infty = \sup_{x \in K} h_k(x) \leq C_H$ for all k .

Substituting this into $\lambda(g_k) = \int_K g_k h_k d\mu$ gives

$$0 \leq \lambda(g_k) \leq \|h_k\|_\infty \int_K g_k d\mu \leq C_H \int_K g_k d\mu.$$

Since $0 \leq g_k \leq M$ and $g_k \rightarrow M\mathbf{1}_F$ pointwise with $\mu(F) = 0$, dominated convergence yields $\int_K g_k d\mu \rightarrow 0$, and hence $\lambda(g_k) \rightarrow 0$.

Finally, by definition of the rate function,

$$I(\nu) \geq \int_K g_k d\nu - \lambda(g_k).$$

Since $g_k \geq M\mathbf{1}_F$ and $\nu(F) = \delta$, we have $\int_K g_k d\nu \geq M\delta$, and letting $k \rightarrow \infty$ gives $I(\nu) \geq M\delta$. As $M > 0$ is arbitrary and $\delta > 0$, we conclude $I(\nu) = \infty$. Therefore $I(\nu) < \infty$ implies $\nu \ll \mu$.

This completes the proof. ■

2.5 Donsker-Varadhan Variational Formula and Decomposition of the Rate Function

This part is inspired by Ferré and Stoltz (2020), Section 3, and we adopt their framework to the setting of a compact domain.

2.5.1 DONSKER-VARADHAN FORMULA FOR $I(\nu)$

Definition 8 (Space $D^+(\mathcal{L}_J)$) *Let*¹

$$D^+(\mathcal{L}_J) = \left\{ u \in D(\mathcal{L}_J) \cap C(K) : u(x) > 0 \text{ for all } x \in K, \text{ and } \frac{\mathcal{L}_J u}{u} \in C_b(K) \right\}.$$

Proposition 9 (Donsker-Varadhan Variational Formula) *For any $\nu \in \mathcal{P}(K)$:*

1. *If $\nu \not\ll \mu$, then $I(\nu) = \infty$. In this case,*

$$\sup_{u \in D^+(\mathcal{L}_J)} \left\{ - \int_K \frac{\mathcal{L}_J u}{u} d\nu \right\} = \infty.$$

2. *If $\nu \ll \mu$, then the rate function $I(\nu)$ can be expressed as:*

$$I(\nu) = \sup_{u \in D^+(\mathcal{L}_J)} \left\{ - \int_K \frac{\mathcal{L}_J u}{u} d\nu \right\}.$$

Proof The proof strategy for showing the equivalence of the Legendre-Fenchel transform $I(\nu)$ and the variational form $I_V(\nu) := \sup_{u \in D^+(\mathcal{L}_J)} \left\{ - \int_K \frac{\mathcal{L}_J u}{u} d\nu \right\}$ largely follows the approach in (Ferré and Stoltz, 2020, Section 6.2.2), adapted to our compact setting.

Part 1: The case when $\nu \not\ll \mu$.

As established in Property (d) of Theorem 7 (LDP for Empirical Measures), if $\nu \not\ll \mu$, then

$$I(\nu) = \sup_{g \in C(K)} \left\{ \int_K g d\nu - \lambda(g) \right\} = \infty.$$

In Step 1 of Part 2 below, we show that for any $\nu \in \mathcal{P}(K)$, it holds that $I_V(\nu) \geq I(\nu)$, where $I_V(\nu) := \sup_{u \in D^+(\mathcal{L}_J)} \left\{ - \int_K \frac{\mathcal{L}_J u}{u} d\nu \right\}$. Specifically, for any $g \in C(K)$, by choosing $u = h_g$ (the principal eigenfunction of $\mathcal{L}_J + g$), we have $u \in D^+(\mathcal{L}_J)$ and

$$- \int_K \frac{\mathcal{L}_J u}{u} d\nu = \int_K g d\nu - \lambda(g).$$

Since this holds for all g , taking the supremum over g yields $I_V(\nu) \geq I(\nu)$.

1. Since K is compact, if $u \in C(K)$, $u > 0$, and $\mathcal{L}_J u \in C(K)$, then $\frac{\mathcal{L}_J u}{u} \in C(K)$, which implies $\frac{\mathcal{L}_J u}{u} \in C_b(K)$.

Therefore, if $\nu \not\ll \mu$, we have $I_V(\nu) \geq I(\nu) = \infty$. This implies $I_V(\nu) = \infty$.

Part 2: The case when $\nu \ll \mu$.

Let

$$I_V(\nu) = \sup_{u \in D^+(\mathcal{L}_J)} \left\{ - \int_K \frac{\mathcal{L}_J u}{u} d\nu \right\}.$$

We aim to show $I(\nu) = I_V(\nu)$.

Step 1: Proof of $I_V(\nu) \geq I(\nu)$. Let $g \in C(K)$ and let $h_g > 0$ be the principal eigenfunction of $\mathcal{L}_J + g$, i.e.

$$(\mathcal{L}_J + g)h_g = \lambda(g)h_g.$$

Dividing by h_g gives

$$-\frac{\mathcal{L}_J h_g}{h_g} = g - \lambda(g).$$

Since $h_g \in D(\mathcal{L}_J) \cap C(K)$ and $g - \lambda(g) \in C(K)$ is bounded on K , we have $h_g \in D^+(\mathcal{L}_J)$. Therefore,

$$I_V(\nu) \geq - \int_K \frac{\mathcal{L}_J h_g}{h_g} d\nu = \int_K (g - \lambda(g)) d\nu = \int_K g d\nu - \lambda(g),$$

where we used $\nu(K) = 1$. Taking the supremum over $g \in C(K)$ yields $I_V(\nu) \geq I(\nu)$.

Step 2: Proof of $I(\nu) \geq I_V(\nu)$. Fix $u \in D^+(\mathcal{L}_J)$ and define

$$f_u := -\frac{\mathcal{L}_J u}{u} \in C(K).$$

Choosing $g = f_u$ in the variational formula for $I(\nu)$ gives

$$I(\nu) \geq \int_K f_u d\nu - \lambda(f_u) = - \int_K \frac{\mathcal{L}_J u}{u} d\nu - \lambda(f_u).$$

It remains to show that $\lambda(f_u) \leq 0$ for all $u \in D^+(\mathcal{L}_J)$. Note that

$$(\mathcal{L}_J + f_u)u = \mathcal{L}_J u + \left(-\frac{\mathcal{L}_J u}{u} \right) u = 0.$$

For $t \geq 0$, the Feynman–Kac formula gives

$$P_t^{f_u} u(x) = \mathbb{E}^x \left[u(X_t) \exp \left(\int_0^t f_u(X_s) ds \right) \right].$$

Applying Itô's formula to $e^{\int_0^t f_u(X_s) ds} u(X_t)$ yields

$$d \left(e^{\int_0^t f_u(X_s) ds} u(X_t) \right) = e^{\int_0^t f_u(X_s) ds} \left((f_u u + \mathcal{L}_J u)(X_t) dt + \sqrt{2} \nabla u(X_t) \cdot dW_t \right).$$

Since $(f_u u + \mathcal{L}_J u) = 0$, the drift vanishes, hence

$$M_t := e^{\int_0^t f_u(X_s) ds} u(X_t)$$

is a local martingale. Because u is continuous and strictly positive on compact K , $0 < \min_K u \leq u \leq \max_K u < \infty$, and f_u is bounded; thus M_t is in fact a martingale and $\mathbb{E}^x[M_t] = M_0 = u(x)$. Consequently,

$$u(x) = \mathbb{E}^x \left[e^{\int_0^t f_u(X_s) ds} u(X_t) \right] \geq \left(\min_{y \in K} u(y) \right) \mathbb{E}^x \left[e^{\int_0^t f_u(X_s) ds} \right].$$

Therefore,

$$\frac{1}{t} \log \mathbb{E}^x \left[e^{\int_0^t f_u(X_s) ds} \right] \leq \frac{1}{t} \log \left(\frac{u(x)}{\min_{y \in K} u(y)} \right) \xrightarrow{t \rightarrow \infty} 0,$$

which implies $\lambda(f_u) \leq 0$. Hence

$$I(\nu) \geq - \int_K \frac{\mathcal{L}_J u}{u} d\nu \quad \text{for all } u \in D^+(\mathcal{L}_J),$$

and taking the supremum over u yields $I(\nu) \geq I_V(\nu)$.

Therefore, for any fixed $u \in D^+(\mathcal{L}_J)$ and $g = f_u = -\frac{\mathcal{L}_J u}{u}$,

$$I(\nu) = \sup_{g \in C(K)} \left\{ \int_K g d\nu - \lambda(g) \right\} \geq \int_K f_u d\nu - \lambda(f_u) \geq - \int_K \frac{\mathcal{L}_J u}{u} d\nu,$$

where the last inequality uses $\lambda(f_u) \leq 0$. Taking the supremum over $u \in D^+(\mathcal{L}_J)$ yields $I(\nu) \geq I_V(\nu)$. Together with Step 1, we conclude $I(\nu) = I_V(\nu)$. \blacksquare

Remark 10 *The definition of $D^+(\mathcal{L}_J)$ requiring $\frac{\mathcal{L}_J u}{u} \in C_b(K)$ is crucial for $f_u = -\frac{\mathcal{L}_J u}{u}$ to be in $C_b(K)$, so that $\lambda(f_u)$ is well-defined within the established framework for $\lambda(g)$ where $g \in C(K)$ (or $C_b(K)$). On a compact domain K , if $u \in C(K)$ and $u > 0$, and $\mathcal{L}_J u \in C(K)$, then $\frac{\mathcal{L}_J u}{u}$ is automatically continuous, and hence bounded, so that the condition simplifies slightly. The core requirement is that u is sufficiently regular for $\mathcal{L}_J u$ to be well-defined and continuous, and u itself is continuous and strictly positive.*

2.5.2 DECOMPOSITION OF $I(\nu)$ INTO SYMMETRIC AND SKEW-SYMMETRIC PARTS

Let $\mu(dx) = \frac{1}{Z} e^{-f(x)} dx$ be the unique invariant probability measure for the process $(X_t)_{t \geq 0}$ in (2.1) under Assumptions 1, 2, and 3, as established in Lemma 1. The infinitesimal generator of $(X_t)_{t \geq 0}$ is given by $\mathcal{L}_J u(x) = \Delta u(x) - \langle \nabla f(x), \nabla u(x) \rangle - \langle J(x) \nabla f(x), \nabla u(x) \rangle$, where the domain $D(\mathcal{L}_J) = \{u \in C^2(K) : \langle \nabla u(x), (I + J(x)) \mathbf{n}(x) \rangle = 0 \text{ on } \partial K\}$ (where $\mathbf{n}(x)$ is the outward unit normal vector) simplifies to functions satisfying Neumann boundary conditions, i.e., $\langle \nabla u(x), \mathbf{n}(x) \rangle = 0$ on ∂K , due to Assumption 2.

To analyze the structure of the rate function $I(\nu)$, we decompose the infinitesimal generator \mathcal{L}_J with respect to the invariant measure μ . Let \mathcal{L}_J^* denote the adjoint of \mathcal{L}_J in the Hilbert space $L^2(K, d\mu)$. The symmetric part \mathcal{L}_S and the skew-symmetric part \mathcal{L}_A of \mathcal{L}_J are defined as:

$$\mathcal{L}_S = \frac{1}{2}(\mathcal{L}_J + \mathcal{L}_J^*), \quad \mathcal{L}_A = \frac{1}{2}(\mathcal{L}_J - \mathcal{L}_J^*). \quad (2.17)$$

Under Assumptions 1 and 2, and using in addition Assumption 3 (the divergence-free condition $\nabla \cdot J = 0$; in particular, since J is skew-symmetric and $f \in C^2$, this implies $\nabla \cdot (J\nabla f) = 0$), the generator admits the symmetric/antisymmetric decomposition $\mathcal{L}_J = \mathcal{L}_S + \mathcal{L}_A$ in $L^2(\mu)$: \mathcal{L}_S is self-adjoint and \mathcal{L}_A is anti-self-adjoint. Moreover, \mathcal{L}_S is self-adjoint, \mathcal{L}_A is anti-self-adjoint with respect to μ , and they have the explicit expressions (see Appendix C for derivation):

$$\mathcal{L}_S u(x) = \Delta u(x) - \langle \nabla f(x), \nabla u(x) \rangle, \quad \mathcal{L}_A u(x) = - \langle J(x) \nabla f(x), \nabla u(x) \rangle. \quad (2.18)$$

Note that \mathcal{L}_S is the infinitesimal generator of the reversible reflected Langevin dynamics (RLD) (1.4). The following proposition details the decomposition of $I(\nu)$ based on these operators.

In the following, we will use the notation $\|\cdot\|_{\mathcal{H}_S^1(\nu)}$ for the Sobolev H^1 -type semi-norm with respect to the measure ν , defined by $\|\varphi\|_{\mathcal{H}_S^1(\nu)}^2 := \int_K \|\nabla \varphi\|_2^2 d\nu$ for a sufficiently regular function φ . The corresponding dual semi-norm, denoted $\|\cdot\|_{\mathcal{H}_S^{-1}(\nu)}$, for an element φ is implicitly defined via a variational problem. Specifically,

$$\|\varphi\|_{\mathcal{H}_S^{-1}(\nu)}^2 = \sup_{\psi \in D^+(\mathcal{L}_J)} \left\{ 2 \int_K \psi \varphi d\nu - \|\psi\|_{\mathcal{H}_S^1(\nu)}^2 \right\}, \quad (2.19)$$

which is consistent with usage in literature such as Ferré and Stoltz (2020). The proposition below details the decomposition of $I(\nu)$ using these concepts.

Proposition 11 (Decomposition of the Rate Function) *Suppose Assumptions 1, 2, and 3 hold. For any $\nu \in \mathcal{P}(K)$ with $\nu \ll \mu$ and $v = \log(d\nu/d\mu)$ satisfying Remark 12, the rate function $I(\nu)$ given by Proposition 9 can be decomposed as: $I(\nu) = I_S(\nu) + I_A(\nu)$, where*

- (a) $I_S(\nu) = \frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu =: \frac{1}{4} \|v\|_{\mathcal{H}_S^1(\nu)}^2$. This term corresponds to the rate function for the reversible reflected Langevin dynamics (RLD) (1.4) governed by \mathcal{L}_S .
- (b) $I_A(\nu) = \frac{1}{4} \|\mathcal{L}_A v\|_{\mathcal{H}_S^{-1}(\nu)}^2$, where $\|\cdot\|_{\mathcal{H}_S^{-1}(\nu)}$ is defined in (2.19).
- (c) The term $I_A(\nu)$ is non-negative, i.e., $I_A(\nu) \geq 0$.

If ν is not absolutely continuous with respect to μ , then $I(\nu) = \infty$. In the absolutely continuous case, the decomposition holds whenever the admissibility conditions in Remark 12 hold; otherwise we set $I(\nu) = \infty$.

Remark 12 (Admissibility and regularity of v) *In Proposition 11, recall that we write $v = \log(d\nu/d\mu)$. We assume that “ v is sufficiently regular”, that is:*

- (i) v belongs to the Sobolev space associated with the symmetric part, i.e.

$$v \in \mathcal{H}_S^1(\nu) \quad \text{if and only if} \quad \int_K \|\nabla v(x)\|_2^2 \nu(dx) < \infty,$$

so that $I_S(\nu) = \frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu$ is well-defined.

(ii) $\mathcal{L}_A v$ belongs to the dual space $\mathcal{H}_S^{-1}(\nu)$, i.e.

$$\mathcal{L}_A v \in \mathcal{H}_S^{-1}(\nu) \quad \text{if and only if} \quad \|\mathcal{L}_A v\|_{\mathcal{H}_S^{-1}(\nu)} < \infty,$$

where $\|\cdot\|_{\mathcal{H}_S^{-1}(\nu)}$ is defined in (2.19). In particular, $\mathcal{L}_A v$ is understood in the weak sense when v is only in $\mathcal{H}_S^1(\nu)$.

If either (i) or (ii) fails (or if $\nu \not\ll \mu$), we set $I(\nu) = \infty$.

Proof of Proposition 11 The proof is inspired from the framework from (Ferré and Stoltz, 2020, Theorem 3.3) for the unconstrained setting. From Proposition 9, the rate function is $I(\nu) = \sup_{u \in D^+(\mathcal{L}_J)} \left\{ - \int_K \frac{(\mathcal{L}_J u)}{u} d\nu \right\}$. Assume $d\nu = e^v d\mu$, where $v = \log(d\nu/d\mu)$. Following Ferré and Stoltz (2020), we choose the test function $u = e^\omega$, where $\omega = v/2 + \psi/2$, and ψ is another test function (assumed to be in $D(\mathcal{L}_J)$ and sufficiently regular). Then, using the fact that the diffusion matrix is the identity,

$$-\frac{(\mathcal{L}_J u)(x)}{u(x)} = -\frac{(\mathcal{L}_J e^\omega)(x)}{e^\omega(x)} = -((\mathcal{L}_J \omega)(x) + \|\nabla \omega(x)\|_2^2).$$

Substituting $\mathcal{L}_J = \mathcal{L}_S + \mathcal{L}_A$ and $\omega = v/2 + \psi/2$, $I(\nu)$ is the supremum over ψ of:

$$\begin{aligned} & - \int_K \left(\mathcal{L}_S \left(\frac{v + \psi}{2} \right) + \mathcal{L}_A \left(\frac{v + \psi}{2} \right) + \left\| \nabla \left(\frac{v + \psi}{2} \right) \right\|_2^2 \right) d\nu \\ &= -\frac{1}{2} \int_K (\mathcal{L}_S v + \mathcal{L}_A v) d\nu - \frac{1}{2} \int_K (\mathcal{L}_S \psi + \mathcal{L}_A \psi) d\nu \\ & \quad - \frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu - \frac{1}{2} \int_K \nabla v \cdot \nabla \psi d\nu - \frac{1}{4} \int_K \|\nabla \psi\|_2^2 d\nu. \quad (2.20) \end{aligned}$$

We use the divergence theorem (integration by parts): for a vector field \mathbf{F} and scalar function g , $\int_K (\nabla \cdot \mathbf{F}) g dx = - \int_K \mathbf{F} \cdot \nabla g dx + \int_{\partial K} g (\mathbf{F} \cdot \mathbf{n}) dS$, where \mathbf{n} is the outward unit normal vector. To simplify terms involving \mathcal{L}_S : $\mathcal{L}_S u = e^f \nabla \cdot (e^{-f} \nabla u)$. The term $-\frac{1}{2} \int_K (\mathcal{L}_S \psi) d\nu$ becomes:

$$\begin{aligned} -\frac{1}{2} \int_K (\mathcal{L}_S \psi) e^v d\mu &= -\frac{1}{2} \int_K \left(e^{f(x)} \nabla \cdot (e^{-f(x)} \nabla \psi(x)) \right) e^{v(x)} \frac{1}{Z} e^{-f(x)} dx \\ &= -\frac{1}{2Z} \int_K \left(\nabla \cdot (e^{-f(x)} \nabla \psi(x)) \right) e^{v(x)} dx \\ &= -\frac{1}{2Z} \left(- \int_K \left(e^{-f(x)} \nabla \psi(x) \right) \cdot (e^{v(x)} \nabla v(x)) dx \right. \\ & \quad \left. + \int_{\partial K} e^{v(x)} \left(e^{-f(x)} \nabla \psi(x) \cdot \mathbf{n}(x) \right) dS \right) \\ &= \frac{1}{2Z} \int_K e^{-f(x)} e^{v(x)} (\nabla \psi(x) \cdot \nabla v(x)) dx \quad (\text{since } \nabla \psi \cdot \mathbf{n} = 0 \text{ on } \partial K) \\ &= \frac{1}{2} \int_K (\nabla \psi \cdot \nabla v) d\nu. \end{aligned}$$

Similarly, $-\frac{1}{2} \int_K (\mathcal{L} S v) d\nu = \frac{1}{2} \int_K \|\nabla v\|_2^2 d\nu$.

For the term $\int_K (\mathcal{L}_A v) d\nu$: Since $\mathcal{L}_A 1 = -\langle J \nabla f, \nabla 1 \rangle = 0$, and \mathcal{L}_A is anti-self-adjoint with respect to μ ($\mathcal{L}_A^* = -\mathcal{L}_A$), we obtain $\int_K (\mathcal{L}_A v) d\nu = \int_K (\mathcal{L}_A v) e^v d\mu$. Use the anti-self-adjointness of \mathcal{L}_A , we can show the above term is zero:

$$\int_K (\mathcal{L}_A v) e^v d\mu = \int_K (\mathcal{L}_A e^v) \cdot 1 d\mu = \int_K e^v (\mathcal{L}_A^* 1) d\mu = - \int_K e^v (\mathcal{L}_A 1) d\mu = 0.$$

Substituting these simplifications into (2.20), the expression becomes:

$$\frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu - \frac{1}{2} \int_K \mathcal{L}_A \psi d\nu - \frac{1}{4} \int_K \|\nabla \psi\|_2^2 d\nu.$$

Using the anti-self-adjointness of \mathcal{L}_A with respect to μ and $\mathcal{L}_A(e^v) = e^v(\mathcal{L}_A v)$:

$$\begin{aligned} -\frac{1}{2} \int_K (\mathcal{L}_A \psi) d\nu &= -\frac{1}{2} \int_K (\mathcal{L}_A \psi) e^v d\mu = \frac{1}{2} \int_K \psi (\mathcal{L}_A e^v) d\mu \\ &= \frac{1}{2} \int_K \psi e^v (\mathcal{L}_A v) d\mu = \frac{1}{2} \int_K (\mathcal{L}_A v) \psi d\nu. \end{aligned}$$

Thus, the expression to maximize over ψ is:

$$\frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu + \frac{1}{2} \int_K (\mathcal{L}_A v) \psi d\nu - \frac{1}{4} \int_K \|\nabla \psi\|_2^2 d\nu.$$

Therefore, the rate function $I(\nu)$ is

$$\begin{aligned} I(\nu) &= \sup_{\psi \in D^+(\mathcal{L}_J)} \left\{ \frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu + \frac{1}{2} \int_K (\mathcal{L}_A v) \psi d\nu - \frac{1}{4} \int_K \|\nabla \psi\|_2^2 d\nu \right\} \\ &= \frac{1}{4} \int_K \|\nabla v\|_2^2 d\nu + \frac{1}{4} \|\mathcal{L}_A v\|_{\mathcal{H}_S^{-1}(\nu)}^2 \\ &= \frac{1}{4} \|v\|_{\mathcal{H}_S^1(\nu)}^2 + \frac{1}{4} \|\mathcal{L}_A v\|_{\mathcal{H}_S^{-1}(\nu)}^2 =: I_S(\nu) + I_A(\nu). \end{aligned}$$

Finally, by the variational representation

$$I_A(\nu) = \sup_{\psi \in D^+(\mathcal{L}_J)} \left\{ \frac{1}{2} \int_K (\mathcal{L}_A v) \psi d\nu - \frac{1}{4} \int_K \|\nabla \psi\|_2^2 d\nu \right\},$$

we immediately have $I_A(\nu) \geq 0$ by choosing $\psi \equiv 0$ (which yields value 0 inside the supremum). This completes the proof. \blacksquare

Let us discuss the implications for acceleration and structural dependence on J . Proposition 11 shows that the rate function $I(\nu)$ for the SRNLD is the sum of the rate function $I_S(\nu)$ for the RLD and a non-negative term $I_A(\nu)$ due to the skew-symmetric part of the dynamics. Thus, $I(\nu) \geq I_S(\nu)$. A larger rate function implies that the probability of observing an empirical measure $\nu \neq \mu$ decays faster as $t \rightarrow \infty$. This indicates that the SRNLD process concentrates more sharply around the target invariant measure μ than the RLD process, thereby providing a theoretical justification for the acceleration observed with non-reversible dynamics.

The term $I_A(\nu) = \frac{1}{4} \|\mathcal{L}_A v\|_{\mathcal{H}_S^{-1}(\nu)}^2$, where $\mathcal{L}_A v(x) = -\langle J(x) \nabla f(x), \nabla v(x) \rangle$ and $v = \log(d\nu/d\mu)$, quantifies this improvement. To understand the intricate structural dependence of $I_A(\nu)$ on $J(x)$ beyond its overall magnitude, we first define the operator $(-\Delta_\nu)^{-1}$. This operator acts on functions $\varphi \in L^2(K, d\nu)$ that have zero mean with respect to the measure ν (i.e., $\int_K \varphi d\nu = 0$). For such a φ , the function $u = (-\Delta_\nu)^{-1} \varphi$ is defined as the unique solution in $H^1(K, d\nu)$ that also has zero mean with respect to ν (i.e., $\int_K u d\nu = 0$) and satisfies the weak formulation of the Neumann problem:

$$\int_K \nabla u(x) \cdot \nabla \chi(x) \nu(dx) = \int_K \varphi(x) \chi(x) \nu(dx), \quad \text{for all } \chi \in H^1(K, d\nu).$$

As established in the derivation of Proposition 11 (specifically, that $\int_K (\mathcal{L}_A v) d\nu = 0$), the term $\mathcal{L}_A v$ meets the zero-mean condition required for $(-\Delta_\nu)^{-1}$ to be well-defined. The rate function component $I_A(\nu)$ can then be expressed using this operator as:

$$I_A(\nu) = \frac{1}{4} \int_K (\mathcal{L}_A v)(x) ((-\Delta_\nu)^{-1}(\mathcal{L}_A v))(x) \nu(dx).$$

If $G_\nu(x, y)$ denotes the Green's function (integral kernel) of the operator $(-\Delta_\nu)^{-1}$, this can be written as a quadratic form:

$$I_A(\nu) = \frac{1}{4} \iint_{K \times K} (-\langle J(x) \nabla f(x), \nabla v(x) \rangle) G_\nu(x, y) (-\langle J(y) \nabla f(y), \nabla v(y) \rangle) \nu(dx) \nu(dy).$$

2.6 Variance Reduction in Skew-Reflected Non-Reversible Langevin Dynamics

In this section, we demonstrate that the non-reversible structure can also lead to a reduction in the asymptotic variance. In the unconstrained setting, it is well-documented that breaking reversibility by adding a suitable skew-symmetric drift can significantly improve sampling efficiency in terms of reducing the asymptotic variance (see e.g. Duncan et al. (2017); Rey-Bellet and Spiliopoulos (2015)). In this section, we will show that this phenomenon also holds in the constrained setting.

Let $g \in C^2(K)$ be an observable, and consider its time-average estimator

$$\hat{g}_t = \frac{1}{t} \int_0^t g(X_s) ds$$

for the true mean $\bar{g} = \int_K g d\mu$. The long-term statistical properties of this estimator are governed by a Central Limit Theorem (CLT). The validity of the CLT for the SRNLD process (2.1) is guaranteed by its strong ergodic properties, which we briefly outline.

The generator \mathcal{L}_J of the SRNLD process (2.1) is a uniformly elliptic operator on a compact, connected domain K with $\partial K \in C^{2,\alpha}$ and reflecting (Neumann) boundary conditions. These conditions imply that the associated Markov semigroup admits a strictly positive and continuous heat kernel (see, e.g., (Davies, 1989, Theorem. 3.3.5)), which ensures that the semigroup is strongly Feller and irreducible on K . Consequently, the process admits a unique invariant probability measure μ (Lemma 1) and is ergodic.

Moreover, the reversible part \mathcal{L}_S (the generator of the reflected Langevin dynamics) has a spectral gap on the mean-zero subspace $L_0^2(\mu)$: there exists $\lambda_P > 0$ (the Poincaré

constant) such that for all $u \in D(\mathcal{L}_S) \cap L_0^2(\mu)$,

$$\langle u, (-\mathcal{L}_S)u \rangle_\mu = \int_K \|\nabla u\|_2^2 d\mu \geq \lambda_P \|u\|_{L^2(\mu)}^2. \quad (2.21)$$

Since Assumptions 2 and 3 ensure \mathcal{L}_A is anti-self-adjoint in $L^2(\mu)$, we have for all $u \in D(\mathcal{L}_J) \cap L_0^2(\mu)$,

$$\operatorname{Re} \langle u, (-\mathcal{L}_J)u \rangle_\mu = \operatorname{Re} \langle u, (-\mathcal{L}_S - \mathcal{L}_A)u \rangle_\mu = \langle u, (-\mathcal{L}_S)u \rangle_\mu \geq \lambda_P \|u\|_{L^2(\mu)}^2. \quad (2.22)$$

In particular, 0 is a simple eigenvalue of \mathcal{L}_J (constants), and $-\mathcal{L}_J$ is injective and coercive on $L_0^2(\mu)$; hence the Poisson equation

$$-\mathcal{L}_J \phi = g, \quad \int_K \phi d\mu = 0, \quad (2.23)$$

has a unique solution $\phi \in D(\mathcal{L}_J) \cap L_0^2(\mu)$ for each $g \in L_0^2(\mu)$, and the inverse $(-\mathcal{L}_J)^{-1}$ is a bounded operator on $L_0^2(\mu)$ with

$$\|(-\mathcal{L}_J)^{-1}g\|_{L^2(\mu)} \leq \lambda_P^{-1} \|g\|_{L^2(\mu)}. \quad (2.24)$$

Therefore, as $t \rightarrow \infty$, the estimator \hat{g}_t satisfies a CLT:

$$\sqrt{t}(\hat{g}_t - \bar{g}) \Rightarrow \mathcal{N}(0, \sigma_{g,J}^2). \quad (2.25)$$

The term $\sigma_{g,J}^2$ is known as the **asymptotic variance**, which quantifies the magnitude of the estimator's statistical fluctuations around the mean. It is formally defined by the time-integrated auto-correlation function:

$$\sigma_{g,J}^2 = 2 \int_0^\infty \mathbb{E}_\mu [(g(X_0) - \bar{g})(g(X_t) - \bar{g})] dt. \quad (2.26)$$

Next lemma is important and its analogous version for the unconstrained setting can be found in Duncan et al. (2016). We conduct our analysis in the Hilbert space of mean-zero functions with respect to the invariant measure μ , defined as:

$$L_0^2(\mu) := \left\{ g \in L^2(K, \mu) : \int_K g(x) d\mu(x) = 0 \right\},$$

equipped with the standard $L^2(K, d\mu)$ inner product $\langle g, h \rangle = \int_K g(x)h(x)d\mu(x)$.

Lemma 13 (Characterizations of Asymptotic Variance) *Let $(X_t)_{t \geq 0}$ be the ergodic process defined by Eq. (2.1) with generator \mathcal{L}_J and invariant measure μ . For any mean-zero observable $g \in L_0^2(\mu)$, the asymptotic variance $\sigma_{g,J}^2$ in the Central Limit Theorem can be characterized by the following equivalent expressions:*

(a) **The Green-Kubo Formula:**

$$\sigma_{g,J}^2 = 2 \int_0^\infty \mathbb{E}_\mu [g(X_0)g(X_t)] dt. \quad (2.27)$$

(b) *The Operator-Theoretic Formula:*

$$\sigma_{g,J}^2 = 2 \langle g, (-\mathcal{L}_J)^{-1} g \rangle_\mu. \quad (2.28)$$

Proof The proof will be provided in Appendix B.5. ■

Remark 14 (PDE Formulation for the Asymptotic Variance) *The asymptotic variance, $\sigma_{g,J}^2$ can be characterized through the solution of a partial differential equation (PDE). The variance is given by the operator-theoretic formula:*

$$\sigma_{g,J}^2 = 2 \langle g, (-\mathcal{L}_J)^{-1} g \rangle_\mu. \quad (2.29)$$

To compute this quantity, one can first find an auxiliary function, $\psi(x)$, by solving for the action of the resolvent operator $(-\mathcal{L}_J)^{-1}$ on the observable $g(x)$:

$$\psi(x) = (-\mathcal{L}_J)^{-1} g(x). \quad (2.30)$$

This is equivalent to solving the following second-order, linear, elliptic partial differential equation for $\psi(x)$:

$$-\mathcal{L}_J \psi(x) = g(x), \quad x \in K. \quad (2.31)$$

By substituting the explicit form of the infinitesimal generator \mathcal{L}_J as defined in the manuscript, we obtain the full PDE:

$$-(\Delta \psi(x) - \langle (I + J(x)) \nabla f(x), \nabla \psi(x) \rangle) = g(x). \quad (2.32)$$

This equation is subject to the boundary conditions imposed by the domain of the operator, $D(\mathcal{L}_J)$. Under Assumption 2 ($J(x)\mathbf{n}(x) = 0$ on ∂K), this simplifies to the standard Neumann boundary condition:

$$\langle \nabla \psi(x), \mathbf{n}(x) \rangle = 0, \quad \text{for every } x \in \partial K, \quad (2.33)$$

where $\mathbf{n}(x)$ is the outward unit normal vector on the boundary of the domain K .

Once the solution $\psi(x)$ to this boundary value problem is found, the asymptotic variance is obtained by computing the inner product with respect to the invariant measure μ :

$$\sigma_{g,J}^2 = 2 \int_K g(x) \psi(x) d\mu(x). \quad (2.34)$$

2.6.1 QUANTITATIVE VARIANCE REDUCTION VIA SPECTRAL ANALYSIS

We now establish that the SRNLD strictly reduces the asymptotic variance compared to the RLD and quantify this reduction explicitly. While the general inequality $\sigma_{g,J}^2 \leq \sigma_{g,0}^2$ can be inferred from abstract operator inequalities, a spectral analysis provides a precise characterization of the efficiency gain.

We work in the Hilbert space $\mathcal{H}^1(\mu)$, defined as the completion of $D(\mathcal{L}_J)$ with respect to the inner product $\langle u, v \rangle_1 := \langle u, (-\mathcal{L}_S)v \rangle_\mu = \int_K \nabla u \cdot \nabla v d\mu$. Let $\mathcal{G} := (-\mathcal{L}_S)^{-1} \mathcal{L}_A$ be the operator characterizing the non-reversible perturbation.

It is important to emphasize that the validity of the following spectral analysis rests fundamentally on Assumption 2 ($J(x)\mathbf{n}(x) = 0$ on ∂K). This condition serves two critical purposes:

1. **Domain Compatibility:** It ensures that the non-reversible generator \mathcal{L}_J shares the same domain (functions satisfying Neumann boundary conditions) as the reversible generator \mathcal{L}_S . This allows for the well-defined operator decomposition $\mathcal{L}_J = \mathcal{L}_S + \mathcal{L}_A$.
2. **Skew-Adjointness:** It eliminates the boundary terms arising from integration by parts (as detailed in Appendix C). Consequently, \mathcal{L}_A becomes anti-self-adjoint on $L^2(\mu)$, which implies that \mathcal{G} is a compact, *skew-adjoint* operator on $\mathcal{H}^1(\mu)$ (i.e., $\mathcal{G}^* = -\mathcal{G}$).

Due to this skew-adjointness, the spectrum of \mathcal{G} consists of purely imaginary eigenvalues $\{\pm i\lambda_n\}_{n \geq 1}$ with $\lambda_n > 0$, and a possible kernel. Let $\{e_n\}_{n \geq 1}$ be the corresponding orthonormal eigenfunctions in $\mathcal{H}^1(\mu)$. Let $\phi_g = (-\mathcal{L}_S)^{-1}g$ be the solution to the Poisson equation for the reversible dynamics, decomposed as $\phi_g = \phi_0 + \sum_{n=1}^{\infty} c_n e_n$, where $\phi_0 \in \ker(\mathcal{G})$ and $c_n = \langle \phi_g, e_n \rangle_1$.

Theorem 15 (Quantitative Variance Reduction) *Under Assumptions 1, 2, and 3, let $\sigma_{g,0}^2$ and $\sigma_{g,J}^2$ be the asymptotic variances for the reversible RLD and non-reversible SRNLD, respectively. The asymptotic variance of SRNLD is explicitly given by:*

$$\sigma_{g,J}^2 = 2\|\phi_0\|_1^2 + 2 \sum_{n=1}^{\infty} \frac{|c_n|^2}{1 + \lambda_n^2}. \quad (2.35)$$

This implies a strict asymptotic variance reduction relative to the reversible case:

$$\sigma_{g,0}^2 - \sigma_{g,J}^2 = 2 \sum_{n=1}^{\infty} |c_n|^2 \frac{\lambda_n^2}{1 + \lambda_n^2} \geq 0. \quad (2.36)$$

The inequality is strict ($\sigma_{g,J}^2 < \sigma_{g,0}^2$) unless the observable g is uncorrelated with the non-reversible part of the dynamics (i.e., $c_n = 0$ for all n).

Proof Recall from Lemma 13 that $\sigma_{g,J}^2 = 2\langle g, (-\mathcal{L}_J)^{-1}g \rangle_\mu$. Using the decomposition $\mathcal{L}_J = \mathcal{L}_S + \mathcal{L}_A$ and the definition of \mathcal{G} , we can factorize the resolvent as $(-\mathcal{L}_J)^{-1} = (I + \mathcal{G})^{-1}(-\mathcal{L}_S)^{-1}$. Substituting $g = (-\mathcal{L}_S)\phi_g$, we obtain:

$$\sigma_{g,J}^2 = 2 \langle (-\mathcal{L}_S)\phi_g, (I + \mathcal{G})^{-1}\phi_g \rangle_\mu = 2 \langle \phi_g, (I + \mathcal{G})^{-1}\phi_g \rangle_1.$$

Since the asymptotic variance is a real scalar, it is determined by the symmetric part of the resolvent operator in $\mathcal{H}^1(\mu)$. Crucially, because Assumption 2 guarantees $\mathcal{G}^* = -\mathcal{G}$, the symmetric part of $(I + \mathcal{G})^{-1}$ is simply $(I - \mathcal{G}^2)^{-1}$. Applying the spectral decomposition of ϕ_g to the operator $(I - \mathcal{G}^2)^{-1}$, where $\mathcal{G}e_n = i\lambda_n e_n$, yields:

$$\langle \phi_g, (I - \mathcal{G}^2)^{-1}\phi_g \rangle_1 = \|\phi_0\|_1^2 + \sum_{n=1}^{\infty} \frac{1}{1 - (i\lambda_n)^2} |c_n|^2 = \|\phi_0\|_1^2 + \sum_{n=1}^{\infty} \frac{|c_n|^2}{1 + \lambda_n^2}.$$

Multiplying by 2 gives Eq. (2.35). For the reversible case ($J \equiv 0$), we have $\mathcal{G} = 0$ (implying $\lambda_n = 0$), which yields $\sigma_{g,0}^2 = 2\|\phi_g\|_1^2 = 2\|\phi_0\|_1^2 + 2 \sum_{n=1}^{\infty} |c_n|^2$. Subtracting the two expressions directly yields Eq. (2.36). Since $\lambda_n^2 > 0$, the difference is non-negative, proving the reduction. The proof is complete. \blacksquare

2.6.2 CONNECTION TO LARGE DEVIATIONS THEORY

We have established a Large Deviation Principle for the empirical measure $L_T = \frac{1}{T} \int_0^T \delta_{X_t} dt$ with a good rate function $I(\nu)$ (Theorem 7). The decomposition $I(\nu) = I_S(\nu) + I_A(\nu)$ immediately suggests that convergence is accelerated, as $I(\nu) \geq I_S(\nu)$. We now show how this LDP structure directly translates to the quantitative variance reduction results obtained in the previous section.

A fundamental result from LDP theory connects the asymptotic variance to the curvature of the rate function for a specific observable. Let

$$I_{g,J}(l) := \inf_{\nu \in \mathcal{P}(K): \int_K g d\nu = l} I(\nu)$$

be the rate function for large deviations of the time-average \hat{g}_t , obtained via the Contraction Principle (see, e.g., (Dembo and Zeitouni, 1998, Theorem 4.2.1)).

Theorem 16 (LDP Formulations for Asymptotic Variance) *Let $\sigma_{g,J}^2$ be the asymptotic variance of the time-average estimator for a mean-zero observable $g \in L_0^2(\mu)$ under the SRNLD process. The following two characterizations of the variance are equivalent:*

(a) *The operator-theoretic formula:*

$$\sigma_{g,J}^2 = 2 \langle g, (-\mathcal{L}_J)^{-1} g \rangle_\mu. \quad (2.37)$$

(b) *The LDP formulation via the curvature of the rate function at the mean ($\bar{g} = 0$):*

$$\sigma_{g,J}^2 = \frac{1}{I''_{g,J}(0)}. \quad (2.38)$$

Proof The proof proceeds by demonstrating that both formulations are equivalent to a third quantity: the second derivative of the Scaled Cumulant Generating Function (SCGF) at the origin.

The connection between the rate function $I_{g,J}(l)$ and the SCGF is established by the Gärtner-Ellis theorem, which provides a variational representation for the rate function. The theorem states that under suitable conditions (finiteness and differentiability of the SCGF, and exponential tightness, which are satisfied on our compact domain K), the rate function is the Legendre-Fenchel transform of the SCGF (see, e.g., (Dembo and Zeitouni, 1998, Theorem 2.3.6)):

$$I_{g,J}(l) = \sup_{\beta \in \mathbb{R}} \{\beta l - \lambda_J(\beta)\}. \quad (2.39)$$

The SCGF $\lambda_J(\beta)$ itself is defined via the long-time limit of moment generating functions:

$$\lambda_J(\beta) = \lim_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E} \left[\exp \left(\beta \int_0^T g(X_t) dt \right) \right]. \quad (2.40)$$

By the Feynman-Kac formula, this limit corresponds to the principal eigenvalue of the perturbed operator $\mathcal{L}_\beta := \mathcal{L}_J + \beta g$. The existence and properties of this eigenvalue for

elliptic operators on compact domains are well-established results in spectral theory, see for example (Evans, 2022, Section 6.5, Theorem 3).

A fundamental property of the Legendre transform is that the second derivatives of a function and its dual are reciprocals at corresponding points. We are interested in the point $l = \bar{g} = 0$. The corresponding dual point is $\beta = 0$, since the derivative of the SCGF at zero is the expectation of the observable:

$$\lambda'_J(0) = \lim_{T \rightarrow \infty} \mathbb{E} \left[\frac{1}{T} \int_0^T g(X_t) dt \right] = \mathbb{E}_\mu[g] = \bar{g} = 0.$$

A key property of the Legendre-Fenchel transform is the duality of derivatives. Let $\beta(l)$ be the value of β that achieves the supremum in the definition of $I_{g,J}(l)$. The first-order optimality conditions imply the dual relations: $I'_{g,J}(l) = \beta(l)$ and $l = \lambda'_J(\beta(l))$. Differentiating the second identity, $l = \lambda'_J(\beta(l))$, with respect to l using the chain rule gives $1 = \lambda''_J(\beta(l)) \cdot \frac{d\beta}{dl}$. Since $\frac{d\beta}{dl} = I''_{g,J}(l)$ from the first identity, we establish the reciprocal relationship between the second derivatives, $I''_{g,J}(l) \cdot \lambda''_J(\beta(l)) = 1$. We are interested in the curvature at the mean value, $l = \bar{g} = 0$. The corresponding dual point is $\beta(0) = 0$, since $\lambda'_J(0) = \mathbb{E}_\mu[g] = 0$. Evaluating at this point yields the desired relation: $I''_{g,J}(0) = \frac{1}{\lambda''_J(0)}$. This shows that the LDP formulation of variance, Eq. (2.38), is equivalent to the statement $\sigma_{g,J}^2 = \lambda''_J(0)$. Now, we show that the operator-theoretic formulation, Eq. (2.37), is also equivalent to $\sigma_{g,J}^2 = \lambda''_J(0)$. The quantity $\lambda''_J(0)$ is the second derivative of the principal eigenvalue of the perturbed operator $\mathcal{L}_\beta = \mathcal{L}_J + \beta g$ at $\beta = 0$. This can be computed using perturbation theory for linear operators. Let $\lambda(\beta)$, u_β , and v_β be the principal eigenvalue, right eigenfunction, and left eigenfunction of \mathcal{L}_β , respectively, with normalizations $\langle u_\beta, v_\beta \rangle_\mu = 1$ and $\langle 1, v_\beta \rangle_\mu = 1$.

The first derivative is given by $\lambda'(0) = \langle g u_0, v_0 \rangle_\mu = \langle g, 1 \rangle_\mu = \bar{g} = 0$, where $u_0 = v_0 = 1$. The second derivative formula, derived from the perturbation series (see, e.g., (Kato, 2013, Chapter II, Equation 2.33)), is given by:

$$\lambda''_J(0) = \langle g u'_0, v_0 \rangle_\mu + \langle g u_0, v'_0 \rangle_\mu, \quad (2.41)$$

where u'_0 and v'_0 are the first-order corrections to the right and left eigenfunctions, respectively. Solving the first-order equations yields $u'_0 = (-\mathcal{L}_J)^{-1}g$ and $v'_0 = (-\mathcal{L}_J^*)^{-1}g$. Due to the duality in $L^2(\mu)$, the two terms in (2.41) are equal:

$$\langle g u_0, v'_0 \rangle_\mu = \langle g, (-\mathcal{L}_J^*)^{-1}g \rangle_\mu = \langle (-\mathcal{L}_J)^{-1}g, g \rangle_\mu = \langle u'_0, g u_0 \rangle_\mu.$$

Thus, the expression simplifies to the dominant term:

$$\lambda''_J(0) = 2 \langle g u'_0, v_0 \rangle_\mu = 2 \langle g (-\mathcal{L}_J)^{-1}g, 1 \rangle_\mu = 2 \langle g, (-\mathcal{L}_J)^{-1}g \rangle_\mu.$$

This formula thus arises fundamentally from solving for the first-order perturbation of the eigenfunction, which necessitates inverting the generator \mathcal{L}_J on the space of mean-zero functions. For a full treatment of perturbation theory, (Kato, 2013, Chapter II) is the standard reference.

By comparing (2.41) with the operator-theoretic variance formula (2.37), we immediately see that, $\sigma_{g,J}^2 = \lambda''_J(0)$. Since we have $\sigma_{g,J}^2 = 1/I''_{g,J}(0) \iff \sigma_{g,J}^2 = \lambda''_J(0)$ and $\sigma_{g,J}^2 = 2 \langle g, (-\mathcal{L}_J)^{-1}g \rangle_\mu \iff \sigma_{g,J}^2 = \lambda''_J(0)$, this establishes the theorem. \blacksquare

Derivation of Quantitative Variance Reduction via LDP. We can now utilize the equivalence $\sigma_{g,J}^2 = 1/I''_{g,J}(0)$ to recover the quantitative spectral formula (2.35) directly from the local geometry of the rate function $I(\nu)$. This confirms that the acceleration phenomenon is encoded in the “steepness” of the LDP rate function.

Consider the tangent space of $\mathcal{P}(K)$ at μ , which can be identified with the space of mean-zero functions $L_0^2(\mu)$. For any direction $v \in L_0^2(\mu)$, let $(\nu_\epsilon)_{\epsilon \geq 0}$ be a smooth path of probability measures starting at μ with initial velocity v (e.g., defined by the density $d\nu_\epsilon \propto e^{\epsilon v} d\mu$). We define the second variation of the rate function along this direction as:

$$\mathcal{Q}(v) := \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon^2} I(\nu_\epsilon).$$

For the symmetric part, a direct calculation using the definition of I_S yields the Dirichlet energy:

$$\mathcal{Q}_S(v) = \frac{1}{4} \int_K \|\nabla v\|_2^2 d\mu = \frac{1}{4} \|v\|_1^2.$$

For the skew-symmetric part, using the operator $\mathcal{G} = (-\mathcal{L}_S)^{-1} \mathcal{L}_A$ and the identity $\mathcal{L}_A = -\mathcal{L}_S \mathcal{G}$, the second variation corresponds to the squared dual norm in $\mathcal{H}^1(\mu)$:

$$\mathcal{Q}_A(v) = \frac{1}{4} \|\mathcal{L}_A v\|_{\mathcal{H}_S^{-1}(\mu)}^2 = \frac{1}{4} \langle \mathcal{L}_A v, (-\mathcal{L}_S)^{-1} \mathcal{L}_A v \rangle_\mu = \frac{1}{4} \|\mathcal{G}v\|_1^2.$$

Thus, the total second variation is strictly equal to the following quadratic form:

$$\mathcal{Q}(v) = \frac{1}{4} (\|v\|_1^2 + \|\mathcal{G}v\|_1^2) = \frac{1}{4} \langle v, (I + \mathcal{G}^* \mathcal{G})v \rangle_1 = \frac{1}{4} \langle v, (I - \mathcal{G}^2)v \rangle_1,$$

where we used the skew-adjointness $\mathcal{G}^* = -\mathcal{G}$ in $\mathcal{H}^1(\mu)$.

To find the variance, we compute the curvature of the observable’s rate function. This amounts to minimizing the quadratic form $\mathcal{Q}(v)$ over all directions v compatible with a small deviation in the observable mean, i.e., subject to the linearized constraint $\langle \phi_g, v \rangle_1 = l$:

$$\frac{1}{2} I''_{g,J}(0) l^2 = \inf_{v \in L_0^2(\mu)} \left\{ \frac{1}{4} \langle v, (I - \mathcal{G}^2)v \rangle_1 \mid \langle \phi_g, v \rangle_1 = l \right\}.$$

The exact solution to this quadratic programming problem is $v^* = \lambda(I - \mathcal{G}^2)^{-1} \phi_g$ for a Lagrange multiplier λ . Substituting this back yields the minimum value:

$$\frac{1}{2} I''_{g,J}(0) l^2 = \frac{l^2}{4 \langle \phi_g, (I - \mathcal{G}^2)^{-1} \phi_g \rangle_1}.$$

Using the relation $\sigma_{g,J}^2 = 1/I''_{g,J}(0)$, we immediately recover the operator-theoretic variance formula:

$$\sigma_{g,J}^2 = 2 \langle \phi_g, (I - \mathcal{G}^2)^{-1} \phi_g \rangle_1.$$

Applying the spectral decomposition of the skew-adjoint operator \mathcal{G} (with eigenvalues $i\lambda_n$) to the term $(I - \mathcal{G}^2)^{-1}$ yields the quantitative series expansion derived in Theorem 15:

$$\sigma_{g,J}^2 = 2 \|\phi_0\|_1^2 + 2 \sum_{n=1}^{\infty} \frac{|c_n|^2}{1 + \lambda_n^2}.$$

This derivation demonstrates that the variance reduction ($\sigma_{g,J}^2 < \sigma_{g,0}^2$) is a direct consequence of the additional positive term $\|\mathcal{G}v\|_1^2$ in the LDP rate function, which increases the curvature $I''_{g,J}(0)$ and thus decreases the variance.

Remark 17 (Connection to the Cramér-Rao Lower Bound) *Note that the identity $\sigma_{g,J}^2 = 1/I''_{g,J}(0)$ in Eq. (2.38) identifies the asymptotic variance with the inverse curvature of the rate function. In the theory of large deviations for estimators, Bahadur (1967) established that this quantity corresponds to the “asymptotic effective variance,” which coincides with the Cramér-Rao lower bound (the inverse Fisher information) for efficient estimators. Thus, $I''_{g,J}(0)$ can be interpreted as the dynamical Fisher information. This suggests that an optimal design of J should aim to maximize the curvature $I''_{g,J}(0)$ to minimize the variance bound.*

3. Numerical Experiments

We present numerical results to validate our theoretical findings, specifically to demonstrate the impact of choosing different skew-symmetric matrices $J(x)$ on the sampling efficiency of *skew-reflected non-reversible Langevin Monte Carlo* (SRNLMC). We compare the performance of standard *projected Langevin Monte Carlo* (PLMC), corresponding to $J(x) \equiv 0$, against SRNLMC employing various forms of $J(x)$. These forms include a constant skew-symmetric matrix and state-dependent skew-symmetric matrices designed such that $J(x)\mathbf{n}(x) = 0$ on the boundary of the unit ball (where $\mathbf{n}(x)$ is the normal vector). Experiments are conducted on a toy problem of sampling a truncated Gaussian, a constrained Bayesian linear regression task, and constrained Bayesian logistic regressions using both synthetic and real data sets. In this work, to make fair comparison, we adopt similar experimental setup as in Du et al. (2025). Our goal is to demonstrate that by using an appropriately chosen state-dependent skew-symmetric matrix $J_s(x)$, SRNLMC can achieve better acceleration performance than simply using a constant skew-symmetric matrix J_a . In particular, consider the constraint set to be a 3-dimensional centered ℓ^2 unit ball such that $K = \{x \in \mathbb{R}^3 : \|x\|_2^2 \leq r\}$, and smoothed ℓ_p ball constraint such that the sublevel domain is $K = \{x \in \mathbb{R}^3 : g(x) \leq \lambda\}$, with $g(x) := \sum_{i=1}^d (x_i^2 + \varepsilon^2)^{\frac{p}{2}}$ as given in (2.5), we will conduct a toy example of truncated multivariate normal distribution, constrained Bayesian linear regression and constrained Bayesian logistic regression using synthetic data. And then we will provide constrained Bayesian logistic regression using real data as an example in high-dimension. In the 3-dimensional examples, we choose the following constant skew-symmetric matrix J_a as a nonstate-dependent skew-symmetric matrix.

$$J_a = \begin{pmatrix} 0 & a & 0 \\ -a & 0 & a \\ 0 & -a & 0 \end{pmatrix}, \quad (3.1)$$

where $a \neq 0$ is a self-chosen constant parameter. In general, we will consider two state-dependent skew-symmetric matrices. For the ℓ^2 unit ball constraint set, the state-dependent

skew-symmetric matrix $J_s(x)$ is defined as the following,

$$J_s(x) = \begin{pmatrix} 0 & -sx_3 & sx_2 \\ sx_3 & 0 & -sx_1 \\ -sx_2 & sx_1 & 0 \end{pmatrix}, \quad (3.2)$$

where $s \neq 0$ is a self-chosen constant parameter. For the smoothed ℓ_p ball constraint with the sublevel-set $\{x \in \mathbb{R}^d : g(x) \leq \lambda\}$, with $g(x) := \sum_{i=1}^d (x_i^2 + \varepsilon^2)^{\frac{p}{2}}$ as given in (2.5), the skew-symmetric matrix $J_g(x)$ is defined as:

$$J_g(x) = \begin{pmatrix} 0 & -k_3^{(a_\ell, b_\ell, c_\ell)}(x) & k_2^{(a_\ell, b_\ell, c_\ell)}(x) \\ k_3^{(a_\ell, b_\ell, c_\ell)}(x) & 0 & -k_1^{(a_\ell, b_\ell, c_\ell)}(x) \\ -k_2^{(a_\ell, b_\ell, c_\ell)}(x) & k_1^{(a_\ell, b_\ell, c_\ell)}(x) & 0 \end{pmatrix}, \quad (3.3)$$

with the coefficients given in (2.13).

In Section 3.1, we will first compare the *iteration complexity* of SRNLMC (J_a) using constant skew-symmetric matrix J_a to the one of SRNLMC ($J_s(x)$), the one of SRNLMC ($J_g(x)$) and the one of *projected Langevin Monte Carlo* (PLMC) in \mathcal{W}_1 distance by using synthetic data in a toy example of the truncated standard multivariate normal distribution. By introducing *stochastic gradients*, we will propose *skew-reflected non-reversible stochastic gradient Langevin dynamics* (SRNSGLD) and *projected stochastic gradient Langevin dynamics* (PSGLD). In Section 3.2, we will compare SRNSGLD to PSGLD in terms of *mean squared error* by using synthetic data in the example of constrained Bayesian linear regression. Furthermore, we will consider the example of constrained Bayesian logistic regression and we compare the *accuracy*, which is defined as the ratio of the correctly predicted labels over the whole dataset in deep learning experiments, of the algorithms by using either synthetic or real data in Section 3.3.

3.1 Toy Example: Truncated Multivariate Normal Distribution

$$\mu(x) = \frac{1}{Z} \exp \left\{ -\frac{x^\top \Sigma^{-1} x}{2} \right\} \mathbf{1}_K, \quad \Sigma = \begin{pmatrix} 0.25 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad (3.4)$$

where $\mathbf{1}_K$ denotes the indicator function on K and $Z := \int_K e^{-\frac{x^\top \Sigma^{-1} x}{2}} dx$ is the normalizing constant.

In our first set of experiments, we take the convex constraint sets to be ball K in \mathbb{R}^3 :

$$K = \{x \in \mathbb{R}^3 : \|x\|_2^2 \leq 1\}, \quad (3.5)$$

we choose the skew-symmetric matrix J_a with $a = 1$ in (3.1) and the state-dependent skew-symmetric matrix J_s with $s = 5$ in (3.2).

In our second set of experiments, we consider the sublevel-set constraint:

$$K = \{x \in \mathbb{R}^3 : g(x) \leq \lambda\}, \quad g(x) := \sum_{i=1}^3 (x_i^2 + \varepsilon^2)^{p/2}. \quad (3.6)$$

that follows from the definition in (2.5). To allow fine-tuning, we incorporate a scaling parameter to the skew-symmetric matrix in (3.3) to obtain:

$$J_g(x) = \begin{pmatrix} 0 & -sk_3^{(a_\ell, b_\ell, c_\ell)}(x) & sk_2^{(a_\ell, b_\ell, c_\ell)}(x) \\ sk_3^{(a_\ell, b_\ell, c_\ell)}(x) & 0 & -sk_1^{(a_\ell, b_\ell, c_\ell)}(x) \\ -sk_2^{(a_\ell, b_\ell, c_\ell)}(x) & sk_1^{(a_\ell, b_\ell, c_\ell)}(x) & 0 \end{pmatrix}, \quad s \neq 0. \quad (3.7)$$

In this toy example, we will take $p = 4$, $\varepsilon = 0.2$ and $\lambda = 1.0$ in (3.6) and take the scaling parameter $s = 8$ in (3.7).

In the proceeding, we simulate 5000 samples and take 2000 iterates starting from the initial point $x_0 = [0.2, 0.3, 0.5]^\top$ with step size $\eta = 5 \times 10^{-4}$, and we first derive a skew-projection formula explicitly by computing the skew unit normal vector direction on the convex set K , and then we approximate the Gibbs distribution by doing rejecting and accepting sampling, i.e. we take *i.i.d.* samples from the 3-dimensional standard normal distribution and discard the samples that are out of the constraint set until obtaining the required number of sample points. Finally, we compare \mathcal{W}_1 distance between the target distribution and the sampling stationary distributions driven by SRNLMC (J_a), SRNLMC ($J_s(x)$), SRNLMC ($J_g(x)$) and PLMC in each dimension.

In Figure 2, we show the first 2 dimensions of the target distribution and the ones of the sampling stationary distributions by SRNLMC (J_a), SRNLMC ($J_s(x)$) and PLMC sequentially. In the same way, we visualize the target distribution and sampling distribution in first 2 dimensions by SRNLMC (J_a), SRNLMC ($J_g(x)$) and PLMC sequentially in Figure 4. These results show that SRNLMC (J_a), SRNLMC ($J_s(x)$), SRNLMC ($J_g(x)$) and PLMC can successfully converge to the truncated standard normal distribution under our setting.

Then we compare the convergence speed of SRNLMC ($J_s(x)$) to either SRNLMC (J_a) or PLMC by computing the \mathcal{W}_1 distance between the target distribution and the sampled stationary distributions in each dimension, and then we do the similar comparison for SRNLMC ($J_g(x)$). We obtain the plots in Figure 3 and Figure 5 where the blue line represents SRNLMC (J_a), the orange line represents SRNLMC ($J_s(x)$) or SRNLMC ($J_g(x)$) and the green line represents PLMC in each subfigure.

Figure 3 and Figure 5 demonstrates that both SRNLMC variants achieve superior iteration complexity compared to PLMC under fixed step sizes, corroborating the theoretical results in Du et al. (2025). Notably, when employing the skew-symmetric matrix $J_s(x)$ (defined in (3.2)) and $J_g(x)$ (defined in (3.3)), both SRNLMC ($J_s(x)$) and SRNLMC ($J_g(x)$) can attain a lower iteration complexity than SRNLMC (J_a) for equivalent \mathcal{W}_1 distance targets.

This empirical finding validates our main result, confirming the theoretical advantage of SRNLMC over PLMC. The enhanced performance stems from the ability of the state-dependent skew-symmetric matrix to provide more effective acceleration in the constrained sampling scenarios compared to its skew-symmetric counterpart J_a . The state-dependent nature of SRNLMC ($J_s(x)$) and SRNLMC ($J_g(x)$) demonstrates robustness improvements over SRNLMC (J_a), where we can observe the orange line leads to better performance. Although constant matrices enforce fixed projection angles, it may fail for skew-projection of the sample points beyond the constraint boundary, the adaptive projection in SRNLMC ($J_s(x)$) and SRNLMC ($J_g(x)$) dynamically adjusts based on the systematical state. This

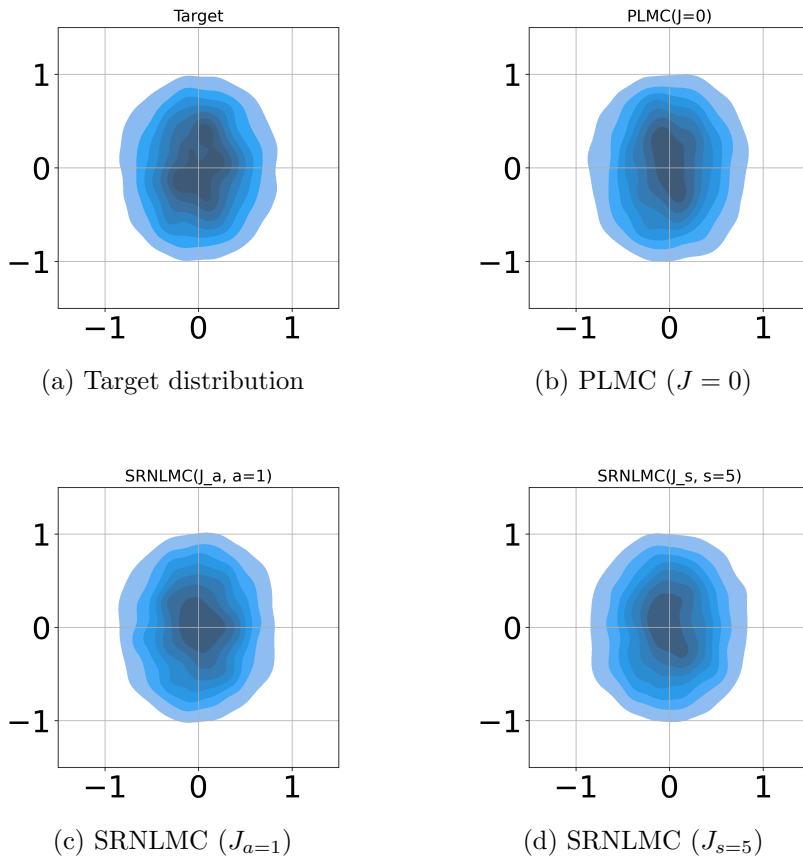


Figure 2: Visualized density plots for the first 2 dimensions with a centered ball constraint.

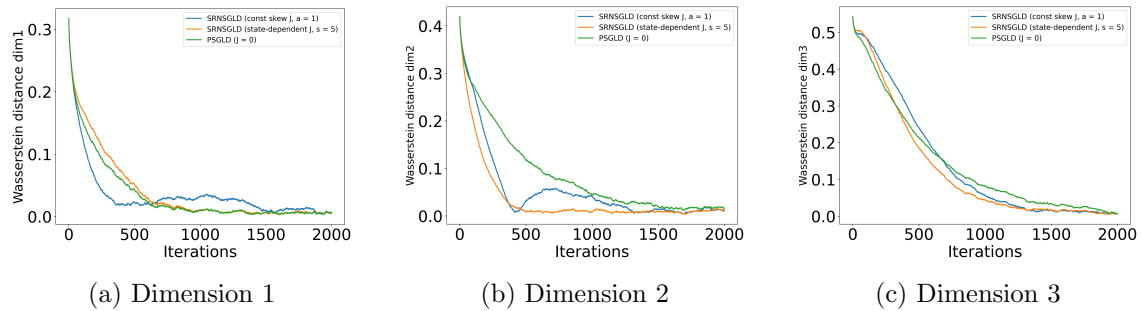


Figure 3: \mathcal{W}_1 distance in each dimension of PLMC and SRNLMLC with a centered ball constraint.

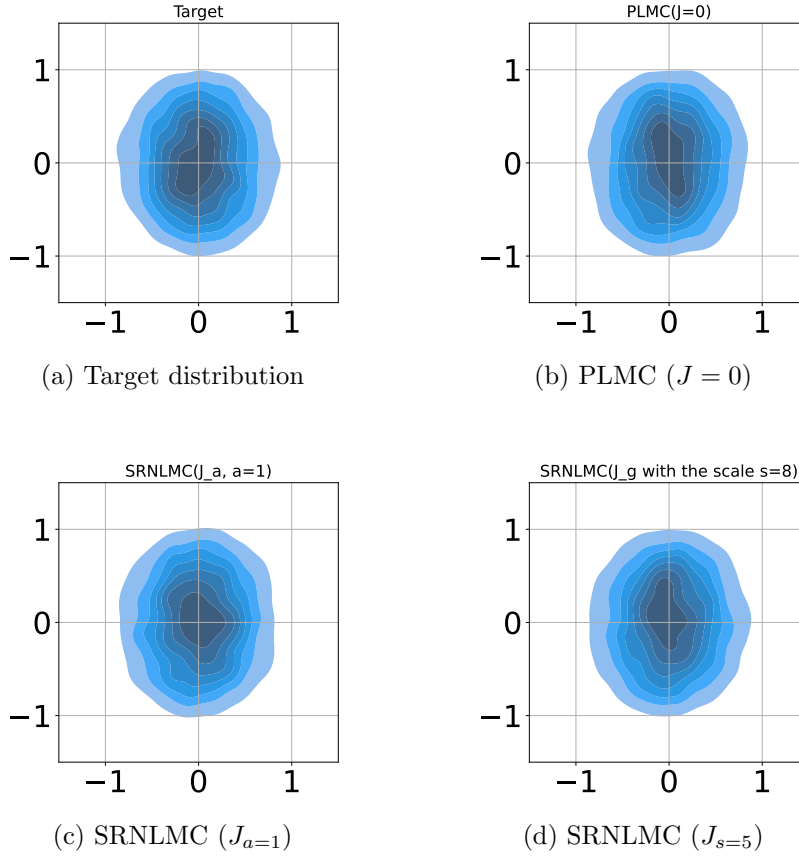


Figure 4: Visualized density plots for the first 2 dimensions with a smoothed ℓ_p ball constraint with a sublevel-set.

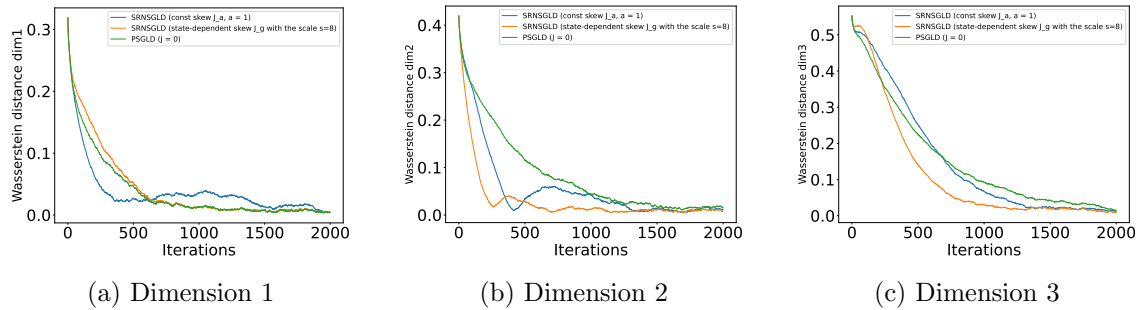


Figure 5: \mathcal{W}_1 distance in each dimension of PLMC and SRNLMC with a smoothed ℓ_p ball constraint with a sublevel-set.

state awareness prevents excessive projection angles to handle extreme sample points and gives room to apply larger step sizes while maintaining equivalent error tolerance levels.

3.2 Constrained Bayesian Linear Regression

In our next set of experiments, we evaluate our algorithms in the context of constrained Bayesian linear regression models. We first consider the example with the ball constraint set, a 3-dimensional centered ℓ^2 unit ball such that $K = \{x \in \mathbb{R}^3 : \|x\|_2^2 \leq 1\}$, which corresponds to the ridge Bayesian linear regression. We consider the linear regression model as the following.

$$\delta_j \sim \mathcal{N}(0, 0.25), \quad a_j \sim \mathcal{N}\left(2 * \mathbf{1}, \frac{1}{2}I_3\right), \quad y_j = x_*^\top a_j + \delta_j, \quad x_* = [1, -0.7, -0.5]^\top, \quad (3.8)$$

where $\mathbf{1}$ denotes the all-one vector. The prior distribution is a uniform distribution, in which case the constraints are satisfied. Our goal is to generate the posterior distribution given by

$$\mu(x) \propto \exp\left\{-\frac{1}{2} \sum_{j=1}^n (y_j - x^\top a_j)^2\right\} \mathbf{1}_K,$$

where $\mathbf{1}_K$ is the indicator function for the constraint set K and n is the total number of data points in the training set.

To further illustrate the efficiency of our algorithms, we will introduce a stochastic gradient setting, and we will show that our algorithms work well in the presence of stochastic gradients. Du et al. (2025) proposed *skew-reflected non-reversible stochastic gradient Langevin dynamics* (SRNSGLD):

$$x_{k+1} = \mathcal{P}_K^J\left(x_k - \eta(I + J(x_k))\nabla f(x_k, \Omega_{k+1}) + \sqrt{2\eta}\xi_{k+1}\right), \quad (3.9)$$

where \mathcal{P}_K^J is the skew-projection onto K , ξ_k are *i.i.d.* Gaussian random vectors $\mathcal{N}(0, I_d)$ and $\nabla f(x_k, \Omega_{k+1})$ is a conditionally unbiased estimator of $\nabla f(x_k)$ over the dataset Ω_{k+1} . For example, a common choice is the mini-batch setting, in which the full gradient is of the form $\nabla f(x) = \frac{1}{n} \sum_{i=1}^n \nabla f_i(x)$, where n is the number of data points and $f_i(x)$ associates with the i -th data point, and the stochastic gradient is given by $\nabla f(x_k, \Omega_{k+1}) := \frac{1}{m} \sum_{i \in \Omega_{k+1}} \nabla f_i(x_k)$ with $m < n$ being the batch-size and $\Omega_{k+1}^{(j)}$ are uniformly sampled random subsets of $\{1, 2, \dots, n\}$ with $|\Omega_{k+1}^{(j)}| = m$ and *i.i.d.* over k . In particular, if $J \equiv 0$, we obtain *projected stochastic gradient Langevin dynamics* (PSGLD).

Under the stochastic gradient setting, we choose the batch size $m = 50$ and generate $n = 10^5$ data points (a_j, y_j) by SRNSGLD, where we take the skew-symmetric matrix J_a with $a = 1$ defined by (3.1) and skew-symmetric matrix $J_s(x)$ with $s = 5$ defined by (3.2). By setting the red starred point as the first 2 dimensions of x_* defined in (3.8) and taking 300 iterates with the step size $\eta = 10^{-4}$, we get Figure 6 and Figure 7 as the following.

Under the same setting, we consider this constrained Bayesian linear regression problem by using a smoothed ℓ_p ball constraint with a sublevel-set defined by $K = \{x \in \mathbb{R}^3 : g(x) \leq \lambda\}$ with $g(x) := \sum_{i=1}^d (x_i^2 + \varepsilon^2)^{p/2}$ from the definition of (2.5), where we take $p = 4$, $\varepsilon = 0.2$,

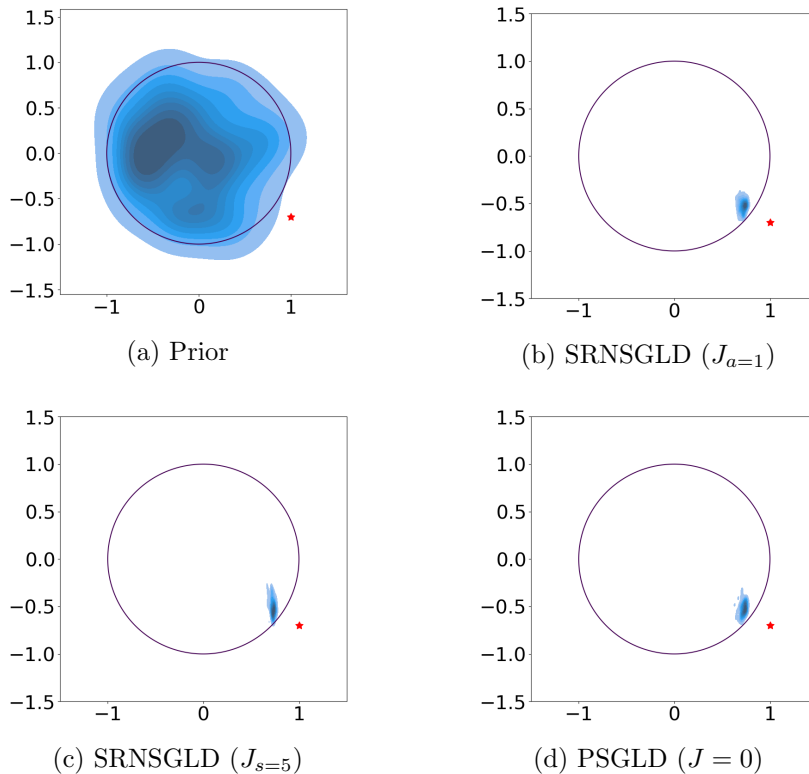


Figure 6: Prior and posterior distributions plots with a centered ball constraint.

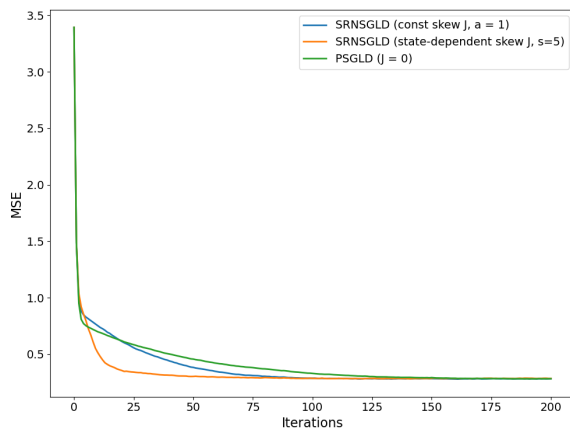


Figure 7: MSE results of SRNSGLD ($J_{a=1}$), SRNSGLD ($J_{s=5}$) and PSGLD ($J = 0$) for the constrained Bayesian linear regression.

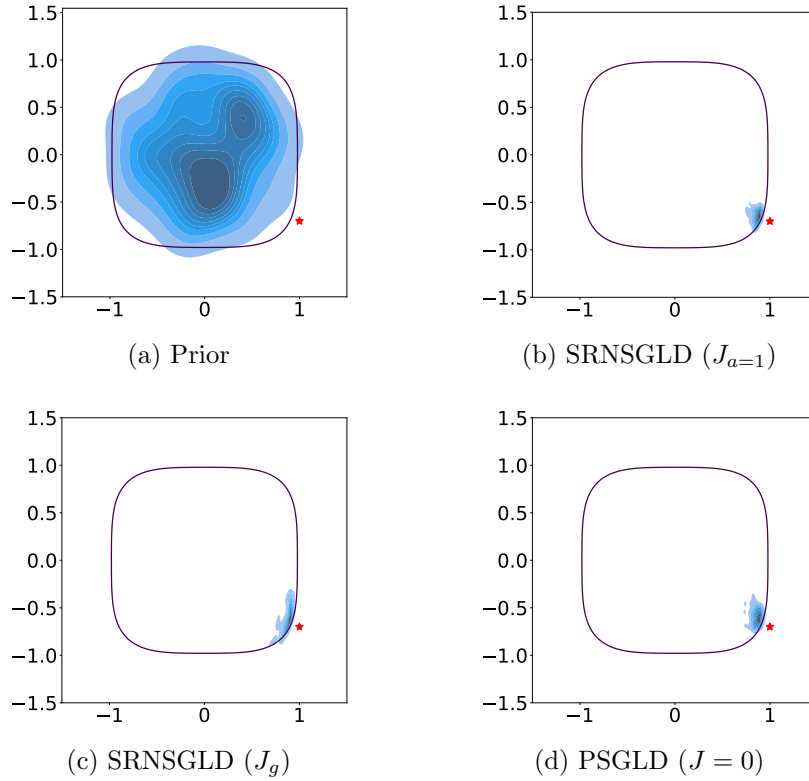


Figure 8: Prior and posterior distributions plots with a smoothed ℓ_p ball constraint with a sublevel-set.

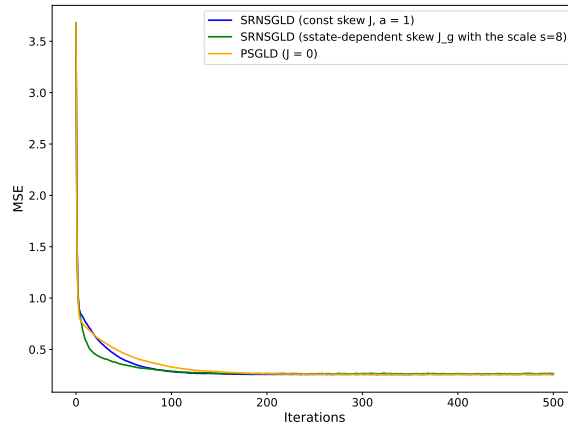


Figure 9: MSE results of SRNSGLD ($J_{a=1}$), SRNSGLD (J_g) and PSGLD ($J = 0$) for the constrained Bayesian linear regression with a smoothed ℓ_p ball constraint with a sublevel-set.

$\lambda = 1.0$ and the scaling parameter $s = 8$ for $J_g(x)$ defined in (3.7). We summarize the results in Figure 8 and Figure 9.

In Figure 6 and Figure 8, the prior uniform distribution is shown in panel (a). In panels (b), (c), (d), our simulation shows that SRNSGLD (J_a), SRNSGLD ($J_s(x)$ in Figure 6), SRNSGLD ($J_g(x)$ in Figure 8) and PSGLD can converge to the samples whose distribution concentrates around the closest position to the target value, which are the red stars shown outside the constraints. By computing the MSE at the k -th run using the formula

$$\text{MSE}_k := \frac{1}{n} \sum_{j=1}^n \left(y_j - (x_k)^\top a_j \right)^2,$$

we get Figure 7 and Figure 9. The blue line represents the mean squared error (MSE) of SRNSGLD (J_a), the orange line corresponds to SRNSGLD ($J_s(x)$) in Figure 7 and SRNSGLD ($J_g(x)$) in Figure 9, and the green line represents PSGLD. Although we do not provide theoretical guarantees for acceleration with respect to MSE, our numerical results show that either SRNSGLD ($J_s(x)$) or SRNSGLD ($J_g(x)$) achieves significantly better acceleration performance than either SRNSGLD (J_a) or PSGLD.

3.3 Constrained Bayesian Logistic Regression

To test the performance of our algorithm in binary classification problems, we implement the constrained Bayesian linear regression models on both synthetic and real data, where the constraint set is the ball of radius r centered at $0 \in \mathbb{R}^d$ given by

$$K_r = \left\{ x \in \mathbb{R}^d : \|x\|_2^2 \leq r \right\}. \quad (3.10)$$

Suppose we can access a dataset $Z = \{z_j\}_{j=1}^n$ where $z_j = (X_j, y_j)$, $X_j \in \mathbb{R}^d$ are the features and $y_j \in \{0, 1\}$ are the labels with the assumption that X_j are independent and the probability distribution of y_j given X_j and the regression coefficients $\beta \in \mathbb{R}^d$ are given by

$$\mathbb{P}(y_j = 1 \mid X_j, \beta) = \frac{1}{1 + e^{-\beta^\top X_j}}.$$

In our experiments, we choose the prior distribution of β to be the uniform distribution in the ball constraint. Then the goal of the constrained Bayesian logistic regression is to sample from $\mu(\beta) \propto e^{-f(\beta)} \mathbf{1}_K$ where:

$$f(\beta) := - \sum_{j=1}^n \log p(y_j \mid X_j, \beta) - \log p(\beta) = \sum_{j=1}^n \log \left(1 + e^{-\beta^\top X_j} \right) + \log V_d, \quad (3.11)$$

and $V_d := \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ is the volume of the unit ball in \mathbb{R}^d .

We will conduct two sets of experiments by using either synthetic data or real data, where one considers the constrained set to be a centered ball defined in (3.10), and another considers the constrained set to be the smoothed ℓ_p ball constraint with a sublevel-set defined by

$$K = \{x \in \mathbb{R}^3 : g(x) \leq \lambda\}, \quad g(x) := \sum_{i=1}^3 (x_i^2 + \varepsilon^2)^{p/2}. \quad (3.12)$$

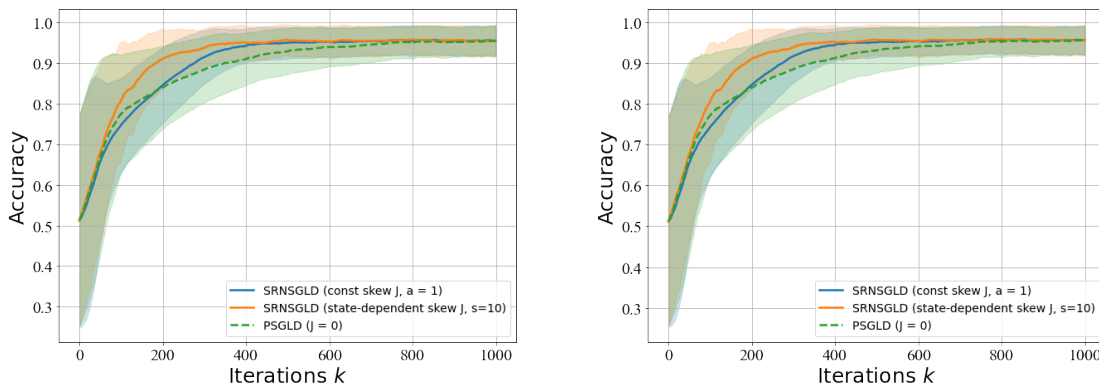
from the definition of (2.5). We will specify the parameters for this constraint set later.

Synthetic Data Consider the following example of $d = 3$. We first generate $n = 2000$ synthetic data by the following model

$$X_j \sim \mathcal{N}(0, 2I_3), \quad p_j \sim \mathcal{U}(0, 1), \quad y_j = \begin{cases} 1 & \text{if } p_j \leq \frac{1}{1+e^{-\beta^\top X_j}}, \\ 0 & \text{otherwise} \end{cases}, \quad (3.13)$$

where $\mathcal{U}(0, 1)$ stands for the uniform distribution on $[0, 1]$ and the prior distribution of $\beta = [\beta_1, \beta_2, \beta_3]^\top \in \mathbb{R}^3$ is a uniform distribution on the centered ball $K_{r=1}$ in \mathbb{R}^3 defined in (3.10). Then we implement them with 1000 iterates by choosing the step size $\eta = 10^{-4}$ with the batch size $m = 50$. For the constrained Bayesian logistic regression, it is not practical for us to compute the \mathcal{W}_1 distance between the approximated Gibbs distribution and the empirical distribution. Hence, for such a binary classification problem, we can use the accuracy over the training set and test set to measure the goodness of the convergence performance.

In the experiment with the ball constraint set in (3.10) with $r = 1$, we choose the skew-symmetric matrix J_a in (3.1) with $a = 1$ for SRNSGLD (J_a) shown by the blue line and the skew-symmetric matrix $J_s(x)$ in (3.2) with $s = 10$ for SRNSGLD ($J_s(x)$) shown by the green dash line. We use 20% of the whole dataset as the test set. The mean and standard deviation of the accuracy distribution are shown in Figure 10.



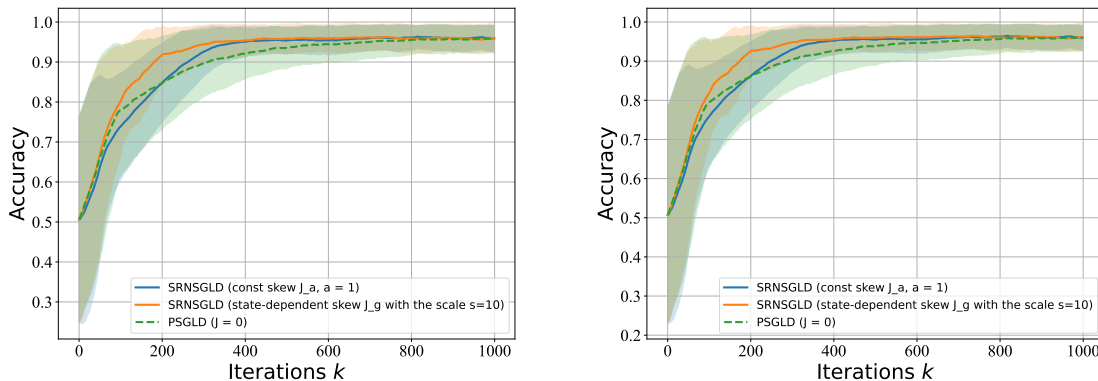
(a) Accuracy over the training set with a centered ball constraint.

(b) Accuracy over the test set with a centered ball constraint.

Figure 10: Accuracy over the training set and the test set for the synthetic data. The blue part (solid line) denotes the mean and standard deviation of SRNSGLD (J_a), the green part (dash-dot line) denotes the mean and standard deviation of SRNSGLD ($J_s(x)$) and the orange part (dashed line) denotes the mean and standard deviation of PSGLD ($J \equiv 0$).

Under the same setting, we consider the smoothed ℓ_p ball constraint with a sublevel-set in (3.12) with $p = 4$ in this synthetic data example, $\varepsilon = 0.2$, $\lambda = 1.0$ and the scaling parameter $s = 10$ for $J_g(x)$ defined in (3.7). We obtain Figure 11.

We observe that, both SRNSGLD ($J_s(x)$) in green line in Figure 10 and SRNSGLD ($J_g(x)$) in orange line in Figure 11 requires fewer iterations to achieve the same level of accuracy on the test set and exhibits more stable sampling behavior comparing to PSGLD in green dash line. Moreover, both SRNSGLD ($J_s(x)$) in green line in Figure 10 and SRNSGLD



(a) Accuracy over the training set with a smoothed ℓ_p ball constraint with a sublevel-set.

(b) Accuracy over the test set with a smoothed ℓ_p ball constraint with a sublevel-set.

Figure 11: Accuracy over the training set and the test set for the synthetic data. The blue part (solid line) denotes the mean and standard deviation of SRNSGLD (J_a), the green part (dash-dot line) denotes the mean and standard deviation of SRNSGLD ($J_g(x)$) and the orange part (dashed line) denotes the mean and standard deviation of PSGLD ($J \equiv 0$).

($J_g(x)$) in orange line in Figure 11 demonstrate superior performance to SRNSGLD (J_a) with constant skew-symmetric matrix and PSGLD ($J = 0$).

Real Data In this section, we consider the constrained Bayesian logistic regression problem on the MAGIC Gamma Telescope dataset² and the Titanic dataset³. The Telescope dataset contains $n = 19020$ samples with dimension $d = 9$, describing the registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. The Titanic dataset contains 891 samples with 10 features representing information about the passengers. And the goal is to predict whether a passenger survived or not based on these features. However, some of the features are irrelevant to our goal, such as the ID number. Therefore, after a pre-proceeding on the raw dataset, we obtained a dataset containing $n = 891$ labeled samples with $d = 9$ features.

For both real data sets, we initialize SRNSGLD and PSGLD with the uniform distribution on the centered unit ball in the respective dimensions. To fit the dimension, we need to adjust the constant skew-symmetric matrix J_a to fit the 9-dimensional setting with the whole superdiagonal set to be a and the subdiagonal set to be $-a$. We let the state-dependent skew-symmetric matrix $J_s(x)$ be a 9×9 skew-symmetric matrix defined for $x \in K_\tau \subset \mathbb{R}^9$. We construct $J_s(x)$ as a block-diagonal matrix:

$$J_s(x) = \begin{pmatrix} J^{(1)}(x_1, x_2, x_3) & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & J^{(2)}(x_4, x_5, x_6) & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & J^{(3)}(x_7, x_8, x_9) \end{pmatrix},$$

2. The Telescope dataset is publicly available from:

<https://archive.ics.uci.edu/ml/datasets/magic+gamma+telescope>.

3. The Titanic dataset is publicly available from:

<https://www.kaggle.com/c/titanic>.

where $\mathbf{0}_{m \times n}$ denotes an $m \times n$ zero matrix and we define the blocks by

$$\begin{aligned} J^{(1)}(x_1, x_2, x_3) &= \begin{pmatrix} 0 & -s_1 x_3 & s_1 x_2 \\ s_1 x_3 & 0 & -s_1 x_1 \\ -s_1 x_2 & s_1 x_1 & 0 \end{pmatrix}, \\ J^{(2)}(x_4, x_5, x_6) &= \begin{pmatrix} 0 & -s_2 x_6 & s_2 x_5 \\ s_2 x_6 & 0 & -s_2 x_4 \\ -s_2 x_5 & s_2 x_4 & 0 \end{pmatrix}, \\ J^{(3)}(x_7, x_8, x_9) &= \begin{pmatrix} 0 & -s_3 x_9 & s_3 x_8 \\ s_3 x_9 & 0 & -s_3 x_7 \\ -s_3 x_8 & s_3 x_7 & 0 \end{pmatrix}. \end{aligned}$$

The 3-dimensional constant $s = [s_1, s_2, s_3]$ can be chosen independently (e.g., $s_1, s_2, s_3 \neq 0$).

Based on the formula of the smoothed ℓ_p ball constraint with a sublevel-set in (3.3), we can also construct the skew-symmetric matrix $J_g(x)$ as a block-diagonal matrix. Moreover, we incorporate a scaling factor $s = [s_1, s_2, s_3] \in \mathbb{R}^3$ into $J_g(x)$ that allows fine-tuning such that

$$J_g(x) = \begin{pmatrix} s_1 J_g^{(1)}(x_1, x_2, x_3) & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & s_2 J_g^{(2)}(x_4, x_5, x_6) & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & s_3 J_g^{(3)}(x_7, x_8, x_9) \end{pmatrix}, \quad (3.14)$$

where $\mathbf{0}_{m \times n}$ denotes an $m \times n$ zero matrix and we denote

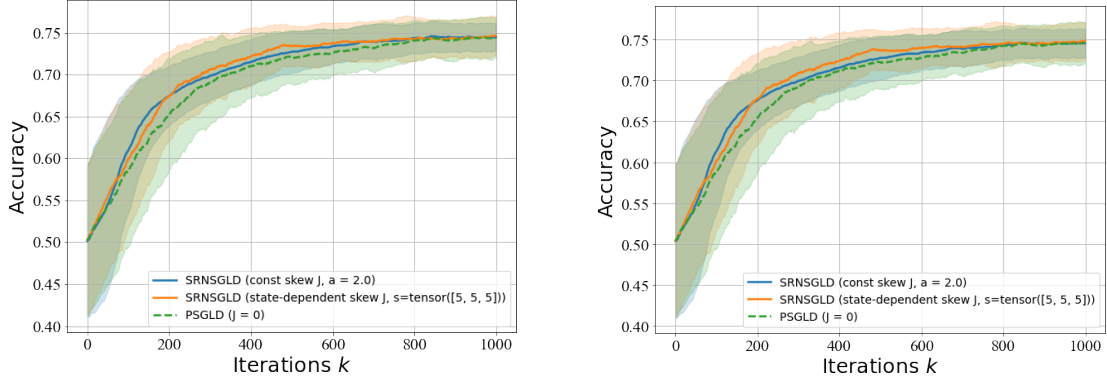
$$J^{(k_1, k_2, k_3)}(x) = \begin{pmatrix} 0 & -k_3^{(a_\ell, b_\ell, c_\ell)}(x) & k_2^{(a_\ell, b_\ell, c_\ell)}(x) \\ k_3^{(a_\ell, b_\ell, c_\ell)}(x) & 0 & -k_1^{(a_\ell, b_\ell, c_\ell)}(x) \\ -k_2^{(a_\ell, b_\ell, c_\ell)}(x) & k_1^{(a_\ell, b_\ell, c_\ell)}(x) & 0 \end{pmatrix},$$

and define

$$\begin{aligned} J_g^{(1)}(x_1, x_2, x_3) &= J^{(k_1, k_2, k_3)}(x), \quad x = [x_1 \ x_2 \ x_3]^\top, \\ J_g^{(2)}(x_4, x_5, x_6) &= J^{(k_1, k_2, k_3)}(x), \quad x = [x_4 \ x_5 \ x_6]^\top, \\ J_g^{(3)}(x_7, x_8, x_9) &= J^{(k_1, k_2, k_3)}(x), \quad x = [x_7 \ x_8 \ x_9]^\top. \end{aligned}$$

In the experiment of using Telescope dataset with a centered ball constraint set, we take the constraint set as the centered ball $K_{r=2} \subset \mathbb{R}^9$ defined in (3.10), and we set the step size $\eta = 10^{-4}$ and batch size $m = 30$ and implement SRNSGLD and PSGLD 1000 iterates with 100 samples over the training set, where we choose the skew-symmetric matrix as J_a defined in (3.1) with $a = 2$ for SRNSGLD (J_a) shown by the blue line and the skew-symmetric matrix as $J_s(x)$ defined in (3.2) with $s = [5, 5, 5]$ for SRNSGLD ($J_s(x)$) shown by the orange dash line. Figure 12 reports the accuracy of two algorithms over the training set and the test set where the test set accounts for 20% of the whole dataset.

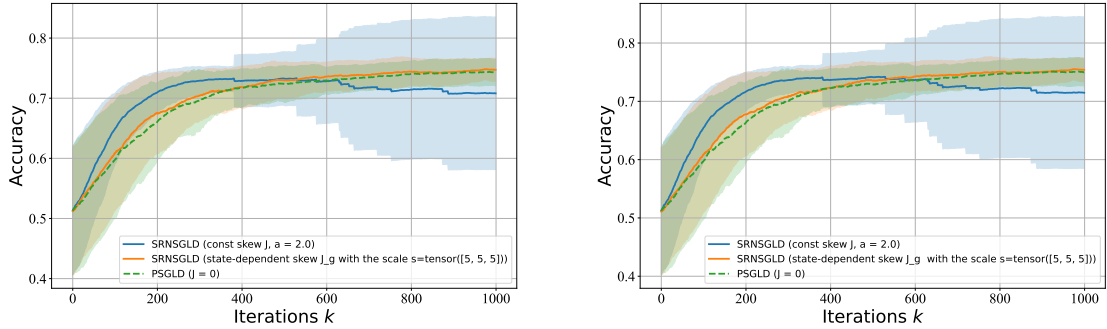
In the experiment of using Telescope dataset with a smoothed ℓ_p ball constraint with a sublevel-set defined in (3.12) with $p = 2.4$ in this Telescope dataset example, $\varepsilon = 0.2$, $\lambda = 4.0$ and the scaling factor of $J_g(x)$ is $s = [5, 5, 5]$ in the definition (3.14). then we



(a) Accuracy over the training set with a centered ball constraint.

(b) Accuracy over the test set with a centered ball constraint.

Figure 12: Accuracy over the training set and the test set for the Telescope dataset. The blue part (solid line) denotes the mean and standard deviation of SRNSGLD (J_a), the orange part (solid line) denotes the mean and standard deviation of SRNSGLD ($J_s(x)$) and the green part (dashed line) denotes the mean and standard deviation of PSGLD ($J \equiv 0$).



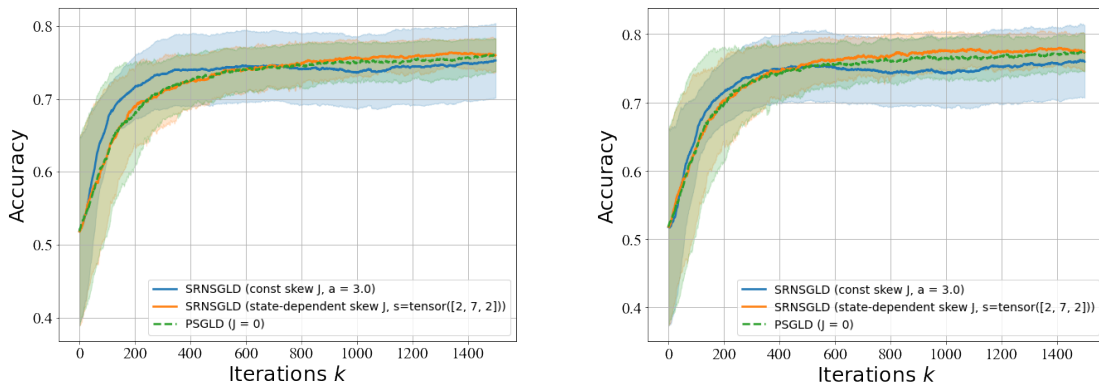
(a) Accuracy over the training set with a smoothed ℓ_p ball constraint with a sublevel-set.

(b) Accuracy over the test set with a smoothed ℓ_p ball constraint with a sublevel-set.

Figure 13: Accuracy over the training set and the test set for the Telescope dataset. The blue part (solid line) denotes the mean and standard deviation of SRNSGLD (J_a), the orange part (solid line) denotes the mean and standard deviation of SRNSGLD ($J_g(x)$) and the green part (dashed line) denotes the mean and standard deviation of PSGLD ($J \equiv 0$).

implement algorithms with the same implementation parameters as the one used in the example with the centered ball constraint set to obtain Figure 13.

For the Titanic data, we first consider the constraint set as the centered ball $K_{r=2} \subset \mathbb{R}^9$, and we set the step size $\eta = 10^{-4}$ and batch size $m = 30$ and implement SRNSGLD and PSGLD 1500 iterates over the training set with 100 samples, we choose the skew-symmetric matrix as J_a defined in (3.1) with $a = 3$ for SRNSGLD (J_a) shown by the green line and the skew-symmetric matrix as $J_s(x)$ defined in (3.2) with $s = [2, 7, 2]$ for SRNSGLD ($J_s(x)$) shown by the orange dash line. Figure 14 reports the accuracy level of algorithms over the training and test sets, where the test set accounts for 20% of the whole dataset.



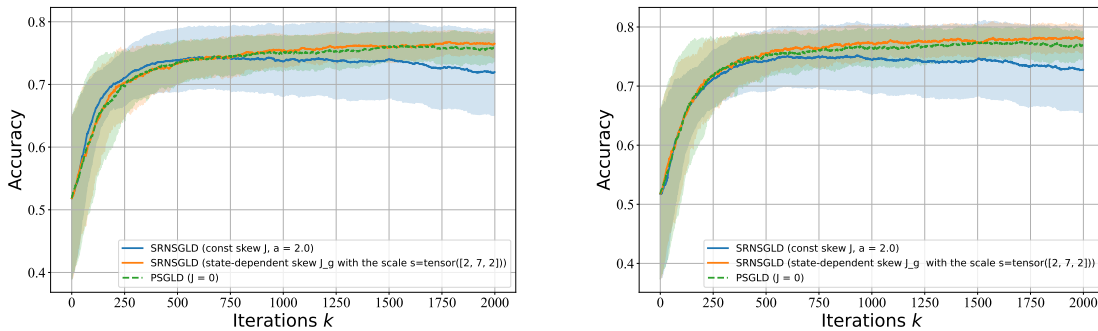
(a) Accuracy over the training set with a centered ball constraint.

(b) Accuracy over the test set with a centered ball constraint.

Figure 14: Accuracy over the training set and the test set for the Titanic dataset. The green part (dash line) denotes the mean and standard deviation of SRNSGLD (J_a), the orange part (dash-dot line) denotes the mean and standard deviation of SRNSGLD ($J_s(x)$) and the blue part (solid line) denotes the mean and standard deviation of PSGLD ($J \equiv 0$).

Taking the constraint set as the smoothed ℓ_p ball constraint with a sublevel-set defined in (3.12) with $p = 2.4$ in this Telescope dataset example, $\varepsilon = 0.18$, $\lambda = 4.0$ and the scaling factor of $J_g(x)$ is $s = [2, 7, 2]$ in the definition (3.14), then we use the same implementation parameters with 2000 steps as for the centered ball constraint set and obtain Figure 15.

From Figure 12 and Figure 13 for Telescope dataset, Figure 14 and Figure 15 for Titanic dataset, we can conclude that both SRNSGLD ($J_s(x)$) and SRNSGLD ($J_g(x)$) consistently outperforms SRNSGLD (J_a) in terms of classification accuracy in the Telescope and Titanic datasets, and the performance of both SRNSGLD ($J_s(x)$) and SRNSGLD ($J_g(x)$) is slightly better than the performance of PSGLD. In particular, SRNSGLD ($J_s(x)$) and SRNSGLD ($J_g(x)$) attain higher accuracy with fewer iterates, which indicates a faster convergence rate. This improvement can be attributed to the adaptive nature and better acceleration rate of the state-dependent skew-symmetric matrix $J_s(x)$ and $J_g(x)$, as it enables more effective exploration of the parameter space compared to the fixed skew-symmetric matrix J_a . In particular, we can observe from Figure 13 and Figure 15 the performance of SRNSGLD (J_a) degrades substantially with the smoothed ℓ_p ball constraint with the sublevel-set.



(a) Accuracy over the training set with a smoothed ℓ_p ball constraint with a sublevel-set.

(b) Accuracy over the test set with a smoothed ℓ_p ball constraint with a sublevel-set.

Figure 15: Accuracy over the training set and the test set for the Titanic dataset. The green part (dash line) denotes the mean and standard deviation of SRNSGLD (J_a), the orange part (dash-dot line) denotes the mean and standard deviation of SRNSGLD ($J_g(x)$) and the blue part (solid line) denotes the mean and standard deviation of PSGLD ($J \equiv 0$).

In summary, these results demonstrate that incorporating a well-chosen state-dependent matrix into the SRNSGLD framework not only leads to theoretical benefits, but also translates into practical efficiency and better empirical performance across different datasets.

4. Concluding Remarks

We established a large deviation principle for empirical measures of skew-reflected non-reversible Langevin dynamics (SRNLD), where a skew-symmetric matrix J with $\nabla \cdot J = 0$ is added to the (reversible) reflected Langevin dynamics (RLD), that samples a target distribution on a constrained domain K . We choose J such that its product with the outward unit normal vector field on the boundary is zero ($J(x)\mathbf{n}(x) = 0$ on ∂K). This simplifies the oblique boundary condition to the standard Neumann condition, which helps overcome the technical challenges that arise from the large deviations theory.

As a by-product, this simplification also provides us a natural choice for the skew-symmetric matrix, that is constructed explicitly for examples including the ball constraint, and more generally the constrained domains characterized by sublevel sets that include special cases such as the smoothed ℓ_p ball constraint. The compactness of K also simplifies several technical aspects, particularly related to the existence of $\lambda(g)$, properties of the Feynman-Kac semigroup, and exponential tightness. The core results on the variational representation and decomposition of the rate function $I(\nu)$ hold, providing valuable insight into the contributions of the reversible (\mathcal{L}_S) and non-reversible (\mathcal{L}_A) parts of the dynamics to the likelihood of fluctuations. The explicit forms for $I_S(\nu)$ and $I_A(\nu)$ depend on the precise definitions of the Sobolev (semi)norms and the properties of \mathcal{L}_A under the invariant measure μ .

By explicitly characterizing the rate functions, we show that SRNLD accelerates the convergence to the target distribution when compared with RLD. Furthermore, we provide

an operator-theoretic proof that these dynamics also reduce the asymptotic variance of time-average estimators, solidifying the advantages of the non-reversible approach.

Numerical experiments based on our choice of the skew-symmetric matrix, that is constructed to satisfy all assumptions in our theory for the ball constraint and the smoothed ℓ_p ball constraint applying to a toy example of truncated multivariate normal distribution, constrained Bayesian linear regression and constrained Bayesian logistic regression using synthetic and real data, show superior performance compared to the existing literature, that validate the theoretical findings from large deviations theory.

There are several promising directions for future research. First, while Theorem 15 provides a spectral characterization of the variance reduction, deriving an explicit *quantitative lower bound* depending solely on the geometry of K and the potential f remains a challenging open problem. Second, building on the connection to the Cramér-Rao lower bound established in Section 2.6.1, future work could focus on the theoretical optimization of the matrix field $J(x)$ to maximize the rate function curvature, thereby minimizing the asymptotic variance. These will be left for future studies.

Acknowledgments

We would like to thank the Action Editor and two anonymous referees whose comments and suggestions have greatly improved the quality of the paper. We also would like to thank Hengrong Du and Qi Feng for helpful discussions. Yingli Wang, Changwei Tu and Xiaoyu Wang are supported by the Guangzhou-HKUST(GZ) Joint Funding Program (No.2025A03J3556). Lingjiong Zhu is partially supported by the grants NSF DMS-2053454 and NSF DMS-2208303.

Appendix A. Summary of Notations

A.1 Definitions and Terminologies

| Term | Description |
|-------------------------------|--|
| $C^k(K)$ | The space of functions defined on K that are k times continuously differentiable. |
| $C^{k,\alpha}(K)$ | The Hölder space consisting of functions $u \in C^k(K)$ whose k -th order partial derivatives are Hölder continuous with exponent $\alpha \in (0, 1]$. The associated norm is $\ u\ _{C^{k,\alpha}(K)} := \sum_{ \beta \leq k} \sup_{x \in K} D^\beta u(x) + \sum_{ \beta =k} \sup_{x,y \in K, x \neq y} \frac{ D^\beta u(x) - D^\beta u(y) }{ x-y ^\alpha}.$ |
| $\partial K \in C^{k,\alpha}$ | The boundary ∂K is of class $C^{k,\alpha}$ if locally it is the graph of a $C^{k,\alpha}$ function. Specifically, for every $x_0 \in \partial K$, there exists a neighborhood U and a bijective map $\psi : U \rightarrow B_1(0)$ such that $\psi, \psi^{-1} \in C^{k,\alpha}$, with $\psi(U \cap K) = \{y \in B_1(0) : y_d \geq 0\}$. |
| Gateaux-differentiable | For a function λ , means that for any $g, h \in C(K)$, the limit $D\lambda(g)[h] := \lim_{\varepsilon \rightarrow 0} \frac{\lambda(g + \varepsilon h) - \lambda(g)}{\varepsilon}$ exists. |
| $H^1(K, d\nu)$ | Sobolev space consisting of functions u defined on domain K such that both u and its first-order weak derivatives are square-integrable with respect to measure ν . More formally, $u \in H^1(K, d\nu)$ if $u \in L^2(K, d\nu)$ and its gradient ∇u (where derivatives are understood in the weak sense) satisfies $\nabla u \in (L^2(K, d\nu))^d$. The squared norm is $\ u\ _{H^1(K, d\nu)}^2 = \int_K u(x) ^2 \nu(dx) + \int_K \ \nabla u(x)\ _2^2 \nu(dx).$ |
| $\mathcal{H}_S^1(\nu)$ | Weak derivatives generalize derivatives to functions that may not be continuously differentiable but still exhibit regularity. A Sobolev H^1 -type semi-norm with respect to measure ν , defined for a sufficiently regular function φ as $\ \varphi\ _{\mathcal{H}_S^1(\nu)}^2 := \int_K \ \nabla \varphi\ _2^2 d\nu.$ |
| $\mathcal{H}_S^{-1}(\nu)$ | This semi-norm is associated with the symmetric part of the infinitesimal generator \mathcal{L}_S . The dual Sobolev semi-norm corresponding to $\ \cdot\ _{\mathcal{H}_S^1(\nu)}$. For an element $\varphi \in L^2(K, d\nu)$ that has zero mean with respect to ν (i.e., $\int_K \varphi d\nu = 0$), it is defined via a variational problem: $\ \varphi\ _{\mathcal{H}_S^{-1}(\nu)}^2 = \sup_{\psi \in D^+(\mathcal{L}_S)} \left\{ 2 \int_K \psi \varphi d\nu - \ \psi\ _{\mathcal{H}_S^1(\nu)}^2 \right\}.$ |

A.2 Sets and Spaces

| Notation | Description |
|----------------------|---|
| \mathbb{R}^d | d -dimensional Euclidean space |
| K | Compact, connected domain with smooth boundary ∂K |
| $C(K)$ | Space of continuous functions on K |
| $C_b(K)$ | Space of bounded continuous functions on K |
| $L^\infty(K)$ | Space of measurable functions on K that are essentially bounded. |
| $\mathcal{P}(K)$ | Space of probability measures on K |
| $D(\mathcal{L}_J)$ | Domain of \mathcal{L}_J , which is $\{u \in C^2(K) : \langle \nabla u(x), (I + J(x))\mathbf{n}(x) \rangle = 0 \text{ on } \partial K\}$ |
| $D^+(\mathcal{L}_J)$ | Space of positive functions u in $D(\mathcal{L}_J) \cap C(K)$ with $\frac{\mathcal{L}_J u}{u} \in C_b(K)$ |

A.3 Operators and Functions

| Notation | Description |
|-----------------|--|
| \mathcal{L}_J | Infinitesimal generator of SRNLD |
| \mathcal{L}_S | Symmetric part of \mathcal{L}_J |
| \mathcal{L}_A | Skew-symmetric part of \mathcal{L}_J |
| ∇ | Gradient operator |
| $\nabla \cdot$ | Divergence operator |
| Δ | Laplacian operator |
| f | Potential function |
| J | Skew-symmetric matrix field |
| \mathbf{n} | outward unit normal vector on ∂K |

A.4 Measures and Norms

| Notation | Description |
|---------------------------------------|---|
| μ | Invariant measure of SRNLD |
| ν | A realization of the empirical measure of X_t |
| $\ \cdot\ _2$ | Euclidean 2-norm |
| $\ \cdot\ _\infty$ | Supremum norm |
| $\ \cdot\ _{\mathcal{H}_S^1(\nu)}$ | Sobolev semi-norm with respect to ν |
| $\ \cdot\ _{\mathcal{H}_S^{-1}(\nu)}$ | Dual Sobolev semi-norm |
| \mathcal{W}_1 | 1-Wasserstein distance |

A.5 Other Notations

| Notation | Description |
|------------------------------------|---|
| $\langle \cdot, \cdot \rangle$ | Inner product in the Euclidean space \mathbb{R}^d |
| $\langle \cdot, \cdot \rangle_\mu$ | Inner product in $L^2(K, d\mu)$ |
| \mathcal{L}_J^* | Adjoint of operator \mathcal{L}_J |
| $I(\nu)$ | Rate function of the empirical measure ν |
| $\lambda(g)$ | Scaled cumulant generating function |

Appendix B. Proofs of Technical Lemmas

B.1 Proof of Lemma 3

1. **Properties of the operator $\mathcal{L}_J + g$ and the semigroup P_t^g :** Since K is a compact domain in \mathbb{R}^d and $g \in C(K)$, g is bounded on K . Let $g_{\max} = \max_{x \in K} |g(x)|$. The operator \mathcal{L}_J of $(X_t)_{t \geq 0}$ in (2.1) is given by: $\mathcal{L}_J u(x) = \Delta u(x) - \langle (I + J(x)) \nabla f(x), \nabla u(x) \rangle$. The term $\Delta u(x)$ is uniformly elliptic. Assume $\partial K \in C^{2,\alpha}$, $f \in C^{2,\alpha}(K)$ and $J \in C^{1,\alpha}(K)$. Then the drift coefficients $(I + J(x)) \nabla f(x)$ belong to $C^{1,\alpha}(K)$, and hence are Hölder continuous. Therefore \mathcal{L}_J is a second-order linear uniformly elliptic operator on K with $C^{1,\alpha}$ coefficients. Moreover, multiplication by $g(x)$ is a bounded (zeroth-order) perturbation (e.g. $g \in L^\infty(K)$). The domain $D(\mathcal{L}_J + g)$ is specified by functions in $C^2(K)$ satisfying Neumann boundary conditions $\langle \nabla u(x), \mathbf{n}(x) \rangle = 0$ on ∂K .

a) **Compactness of P_t^g :** The semigroup P_t^g is generated by $\mathcal{L}_J + g$. For an elliptic operator on a compact manifold with boundary conditions such as Neumann, the resolvent $(\zeta I - (\mathcal{L}_J + g))^{-1}$ is compact for ζ in the resolvent set. This implies that the semigroup $P_t^g = e^{t(\mathcal{L}_J + g)}$ is compact for all $t > 0$ (see, e.g., (Pazy, 1983, Chapter 2, Corollary 3.5)). Alternatively, P_t^g can be represented via the fundamental solution (Green's function) $p^g(t, x, y)$ of the parabolic PDE $\partial_t u = (\mathcal{L}_J + g)u$. For elliptic operators, this fundamental solution is smooth for $t > 0$ (see, e.g., (Stroock, 2008, Chapter 3.3, Theorem 3.3.11)). Since K is compact, $(P_t^g \phi)(x) = \int_K p^g(t, x, y) \phi(y) dy$ is an integral operator with a smooth kernel, which implies that P_t^g is compact on $C(K)$ (also on $L^p(K)$).

b) **Strong positivity of P_t^g .**

Recall that an operator T on $C(K)$ is *strongly positive* if for any $\phi \in C(K)$ with $\phi \geq 0$ and $\phi \not\equiv 0$, one has $T\phi(x) > 0$ for all $x \in K$. Let $(P_t)_{t \geq 0}$ be the Markov semigroup associated with the reflected diffusion generated by \mathcal{L}_J on the bounded connected domain K . Since the diffusion matrix is the identity, \mathcal{L}_J is uniformly elliptic; in particular the Neumann heat kernel $p(t, x, y)$ exists, is jointly continuous on $(0, \infty) \times K \times K$, and is *strictly positive* for every $t > 0$ and $x, y \in K$ (see, e.g., (Davies, 1989, p. 98)). Consequently, for any $\phi \geq 0$, $\phi \not\equiv 0$,

$$(P_t \phi)(x) = \int_K p(t, x, y) \phi(y) dy > 0, \quad x \in K,$$

so P_t is positivity improving; equivalently, the semigroup is irreducible in the sense of (Davies, 1989, Theorem 3.3.5).

Now let $g \in L^\infty(K)$ and define the Feynman–Kac semigroup

$$(P_t^g \phi)(x) := \mathbb{E}^x \left[\phi(X_t) \exp \left(\int_0^t g(X_s) ds \right) \right].$$

The exponential weight is strictly positive and satisfies

$$e^{-t\|g\|_\infty} \leq \exp \left(\int_0^t g(X_s) ds \right) \leq e^{t\|g\|_\infty}.$$

Hence, for $\phi \geq 0$,

$$(P_t^g \phi)(x) \geq e^{-t\|g\|_\infty} \mathbb{E}^x[\phi(X_t)] = e^{-t\|g\|_\infty} (P_t \phi)(x) > 0, \quad x \in K,$$

which proves that P_t^g is strongly positive for every $t > 0$.

2. Existence of a principal eigenvalue/eigenfunction and t -consistency:

Fix some $t_0 > 0$. Since $P_{t_0}^g$ is compact and strongly positive on the Banach lattice $C(K)$, the Krein–Rutman theorem (see, e.g., (Deimling, 2013, Chapter 6, Theorem 19.2)) yields that the spectral radius $r(P_{t_0}^g) > 0$ is a simple eigenvalue of $P_{t_0}^g$ and there exists $h_g \in C(K)$ with $h_g > 0$ on K such that

$$P_{t_0}^g h_g = r(P_{t_0}^g) h_g.$$

Write $r(P_{t_0}^g) = e^{t_0 \lambda(g)}$ and define $\lambda(g) := \frac{1}{t_0} \log r(P_{t_0}^g)$.

Step 1: the same h_g works for all $t \geq 0$. For any $t \geq 0$, by the semigroup property,

$$P_{t_0}^g (P_t^g h_g) = P_t^g (P_{t_0}^g h_g) = r(P_{t_0}^g) P_t^g h_g.$$

Thus $P_t^g h_g$ is a (strictly positive) eigenfunction of $P_{t_0}^g$ associated with the simple eigenvalue $r(P_{t_0}^g)$. By uniqueness up to scaling, there exists a scalar $c(t) > 0$ such that

$$P_t^g h_g = c(t) h_g \quad (t \geq 0).$$

Again by the semigroup property, $c(t+s) = c(t)c(s)$ and $c(0) = 1$. Since $(P_t^g)_{t \geq 0}$ is strongly continuous on $C(K)$ (as X_t is Feller and g is bounded), $t \mapsto P_t^g h_g$ is continuous in $C(K)$. Hence $c(t)$ is continuous, and $c(t) = e^{t\lambda(g)}$ for all $t \geq 0$. Therefore,

$$P_t^g h_g = e^{t\lambda(g)} h_g, \quad \text{for any } t \geq 0.$$

In particular, for every fixed $t > 0$, the principal eigenfunction of P_t^g is unique up to normalization; thus it coincides with h_g after rescaling and the corresponding exponent satisfies $\lambda_t(g) = \lambda(g)$.

Step 2: eigenvalue problem for the generator. Since $t_0 > 0$ and $\mathcal{L}_J + g$ is uniformly elliptic with smooth coefficients and Neumann boundary condition, the kernel $p^g(t_0, x, y)$ is smooth in (x, y) . Hence $P_{t_0}^g$ is smoothing, and in particular $P_{t_0}^g h_g \in D(\mathcal{L}_J + g)$ (see, e.g., (Stroock, 2008, Chapter 3.3, Theorem 3.3.11)). From $P_{t_0}^g h_g = e^{t_0 \lambda(g)} h_g$ we deduce $h_g \in D(\mathcal{L}_J + g)$. From $P_t^g h_g = e^{t\lambda(g)} h_g$ and the definition of the generator,

$$(\mathcal{L}_J + g)h_g = \lim_{t \downarrow 0} \frac{P_t^g h_g - h_g}{t} = \lim_{t \downarrow 0} \frac{e^{t\lambda(g)} - 1}{t} h_g = \lambda(g) h_g,$$

so that h_g is a (strictly positive) eigenfunction of $\mathcal{L}_J + g$ with eigenvalue $\lambda(g)$.

3. Existence of the limit and independence of x : From the definition of the principal eigenvalue, for the strictly positive eigenfunction h_g , we have $(P_t^g h_g)(x) = e^{\lambda(g)t} h_g(x)$. This can be written as $\mathbb{E}^x \left[h_g(X_t) \exp \left(\int_0^t g(X_s) ds \right) \right] = e^{\lambda(g)t} h_g(x)$. Thus,

$\frac{1}{t} \log \mathbb{E}^x \left[\frac{h_g(X_t)}{h_g(x)} \exp \left(\int_0^t g(X_s) ds \right) \right] = \lambda(g)$. Since K is compact and $h_g \in C(K)$ with $h_g(x) > 0$, there exist $0 < c_1 < c_2 < \infty$ such that $c_1 \leq h_g(x) \leq c_2$ for all $x \in K$. Therefore, $\frac{c_1}{c_2} \leq \frac{h_g(X_t)}{h_g(x)} \leq \frac{c_2}{c_1}$. Then,

$$\lambda(g) + \frac{1}{t} \log \left(\frac{c_1}{c_2} \right) \leq \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g(X_s) ds \right) \right] \leq \lambda(g) + \frac{1}{t} \log \left(\frac{c_2}{c_1} \right). \quad (\text{B.1})$$

Taking the limit as $t \rightarrow \infty$ in (B.1), we get $\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g(X_s) ds \right) \right] = \lambda(g)$. This limit exists, is equal to the principal eigenvalue of $\mathcal{L}_J + g$, and is independent of the initial state $x \in K$. The finiteness of $\lambda(g)$ follows because g is bounded on the compact set K . If $g_{\min} \leq g(x) \leq g_{\max}$, then

$$\begin{aligned} g_{\min} &= \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g_{\min} ds \right) \right] \\ &\leq \lambda(g) \leq \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}^x \left[\exp \left(\int_0^t g_{\max} ds \right) \right] = g_{\max}. \end{aligned}$$

Thus, $\lambda(g)$ is finite.

The results cited (e.g., Krein-Rutman theorem (Deimling, 2013, Chapter 6, Theorem 19.2), (Pazy, 1983, Chapter 2, Corollary 3.5)) are standard tools for analyzing such semigroups on compact domains. Donsker and Varadhan's original work on LDPs for Markov processes on compact spaces also establishes the existence of $\lambda(g)$ as the principal eigenvalue (Donsker and Varadhan, 1975a). This completes the proof.

B.2 Proof of Lemma 4

This is a standard consequence of Hölder's inequality. For any $\alpha \in [0, 1]$, $g_1, g_2 \in C(K)$,

$$\begin{aligned} \mathbb{E} \left[e^{\int_0^t (\alpha g_1(X_s) + (1-\alpha)g_2(X_s)) ds} \right] &= \mathbb{E} \left[\left(e^{\int_0^t g_1(X_s) ds} \right)^\alpha \left(e^{\int_0^t g_2(X_s) ds} \right)^{1-\alpha} \right] \\ &\leq \left(\mathbb{E} \left[e^{\int_0^t g_1(X_s) ds} \right] \right)^\alpha \left(\mathbb{E} \left[e^{\int_0^t g_2(X_s) ds} \right] \right)^{1-\alpha}. \end{aligned} \quad (\text{B.2})$$

Taking $\frac{1}{t} \log$ and letting $t \rightarrow \infty$ in (B.2) yields: $\lambda(\alpha g_1 + (1-\alpha)g_2) \leq \alpha \lambda(g_1) + (1-\alpha) \lambda(g_2)$. This completes the proof.

B.3 Proof of Lemma 5

The proof relies on perturbation theory for linear operators, specifically for isolated simple eigenvalues. Let $T(g) = \mathcal{L}_J + M_g$, where M_g is the multiplication operator by g . From the previous proposition (existence of $\lambda(g)$), we know that for each $g \in C(K)$, $\lambda(g)$ is a simple, isolated eigenvalue of $T(g)$ with the largest real part. Let $h_g \in C(K)$ be a corresponding strictly positive eigenfunction, and let h_g^* be a corresponding strictly positive eigenfunction of the adjoint operator $T(g)^* = \mathcal{L}_J^* + M_g$ (adjoint with respect to the $L^2(K, dx)$ inner product, assuming appropriate domains). Since h_g and h_g^* are unique up to positive multiplicative constants, we can normalize them such that their $L^2(K, dx)$ inner product

is $\langle h_g, h_g^* \rangle_{L^2(K, dx)} = \int_K h_g(x) h_g^*(x) dx = 1$. Note that since K is compact and $h_g \in C(K)$ with $h_g(x) > 0$, h_g is bounded from above and below by positive constants on K .

1. Continuity of $\lambda(g)$: We have already established in Proposition 4 that $\lambda(g)$ is convex. A convex function on a Banach space (such as $C(K)$) is continuous if it is bounded on some open set (see, e.g., (Ekeland and Temam, 1999, Chapter 1, Lemma 2.1)). Since $g_{\min} \leq \lambda(g) \leq g_{\max}$ (where $g_{\min} = \inf_K g(x)$ and $g_{\max} = \sup_K g(x)$), if g is in an open ball of radius R around 0 in $C(K)$, then $\|g\|_\infty \leq R$, which implies $-R \leq \lambda(g) \leq R$. Thus $\lambda(g)$ is bounded on open sets. Being convex and bounded on open sets, $\lambda(g)$ is continuous.

2. Gateaux-Differentiability of $\lambda(g)$: We want to show that for any $g, h \in C(K)$, the limit $D\lambda(g)[h] = \lim_{\varepsilon \rightarrow 0} \frac{\lambda(g+\varepsilon h) - \lambda(g)}{\varepsilon}$ exists. Since $\lambda(g)$ is a simple eigenvalue of $T(g) = \mathcal{L}_J + M_g$, its first-order variation with respect to a linear perturbation $T(g) + \varepsilon M_h$ is given by standard perturbation theory (see, e.g., (Kato, 2013, Chapter II, Equation (2.33)) for the coefficient of the first-order term in the eigenvalue series expansion). Let $T(\varepsilon) = T(g + \varepsilon h)$. The eigenvalue is $\lambda(\varepsilon) = \lambda(g + \varepsilon h)$ and the corresponding eigenfunction is $h(\varepsilon) = h_{g+\varepsilon h}$. The differentiability of $\lambda(\varepsilon)$ and $h(\varepsilon)$ with respect to ε at $\varepsilon = 0$ is assured for such analytic perturbations (see (Kato, 2013, Chapter VII, Theorem 2.6, Theorem 3.6)).

The first-order correction to $\lambda(g)$, denoted $\lambda'(0) = \left. \frac{d\lambda(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0}$, can be found by differentiating the eigenvalue equation $T(\varepsilon)h(\varepsilon) = \lambda(\varepsilon)h(\varepsilon)$ with respect to ε at $\varepsilon = 0$:

$$M_h h_g + T(g)h'(0) = \lambda'(0)h_g + \lambda(g)h'(0),$$

where $h_g = h(0)$ is the eigenfunction for $T(g)$. Taking the inner product with the eigenfunction h_g^* of the adjoint $T(g)^*$ (normalized so that $\langle h_g, h_g^* \rangle_{L^2(K, dx)} = 1$), and using $T(g)^*h_g^* = \lambda(g)h_g^*$, we obtain: $\langle M_h h_g, h_g^* \rangle_{L^2(K, dx)} = \lambda'(0)$. Thus, $\lambda'(0) = \int_K h(x) h_g(x) h_g^*(x) dx$. This shows that the Gateaux derivative $D\lambda(g)[h]$ exists and is given by $D\lambda(g)[h] = \int_K h(y) \nu_g(dy)$, where the measure ν_g is defined by

$$\nu_g(dy) = h_g(y) h_g^*(y) dy, \quad \text{with} \quad \int_K h_g(y) h_g^*(y) dy = 1.$$

In particular, ν_g is a probability measure, consistent with the eigenfunction representation used in the main text.

The differentiability of $\lambda(g + \varepsilon h)$ with respect to ε (and in fact, its analyticity for real ε) is a deep result from the perturbation theory of operators. The family of operators $T(g) = \mathcal{L}_J + M_g$ is an affine (and thus analytic) family of operators of type (A) in the sense of (Kato, 2013, Chapter VII, Section 2.1), when considered as operators from, say, $H^2(K) \cap \{\text{BCs}\}$ to $L^2(K)$. Since $\lambda(g)$ is a simple isolated eigenvalue, it is an analytic function of g (if $C(K)$ is considered as a real Banach space, $\lambda(g)$ is real-analytic). Real analyticity implies Fréchet differentiability, which in turn implies Gateaux differentiability. The proof is complete.

B.4 Proof of Lemma 6

By definition (see, e.g., (Dembo and Zeitouni, 1998, Definition 1.2.18)), a family of probability measures $(\mathbb{Q}_t)_{t \geq 0}$ on a Polish space \mathcal{X} is exponentially tight if for every $A < \infty$, there exists a compact set $\Gamma_A \subset \mathcal{X}$ such that $\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{Q}_t(\Gamma_A^c) \leq -A$. In our case, the

Polish space is $\mathcal{X} = \mathcal{P}(K)$, the space of probability measures on K , equipped with the weak topology. The measures are $\mathbb{Q}_t(\cdot) = \mathbb{P}(L_t \in \cdot)$.

Since K is a compact metric space, by Prokhorov's theorem (see, e.g., (Billingsley, 2013, Theorem 5.1 and Theorem 5.2)), the space $\mathcal{P}(K)$ itself is compact in the weak topology. Now, to demonstrate exponential tightness, for any given $A < \infty$, we can choose the compact set $\Gamma_A = \mathcal{P}(K)$. Since L_t is, by its definition as an empirical measure, always an element of $\mathcal{P}(K)$ (it is a probability measure on K), the event $L_t \notin \Gamma_A = \mathcal{P}(K)$ is an impossible event. Therefore, $L_t \notin \mathcal{P}(K)$ implies $L_t \in \emptyset$. Thus, $\mathbb{P}(L_t \notin \Gamma_A) = \mathbb{P}(L_t \notin \mathcal{P}(K)) = \mathbb{P}(L_t \in \emptyset) = 0$. Then, for any $t > 0$, $\frac{1}{t} \log \mathbb{P}(L_t \notin \Gamma_A) = \frac{1}{t} \log 0$. The logarithm of zero is formally $-\infty$. So, $\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}(L_t \notin \Gamma_A) = \limsup_{t \rightarrow \infty} (-\infty) = -\infty$. Since for any $A < \infty$, we have $-\infty \leq -A$, the condition for exponential tightness is satisfied.

This argument is significantly simpler than the one required for non-compact state spaces (such as (Ferré and Stoltz, 2020, Lemma 6.9)), where one typically needs to construct a family of precompact sets using a Lyapunov function Ψ and show that the probability of the empirical measure lying outside these precompact sets decays sufficiently fast. The compactness of the entire space $\mathcal{P}(K)$ makes this step trivial. The proof is complete.

B.5 Proof of Lemma 13

The two formulas are classical and equivalent characterizations of the asymptotic variance for an ergodic Markov process.

(a) The first formula, $\sigma_{g,J}^2 = 2 \int_0^\infty \mathbb{E}_\mu[g(X_0)g(X_t)]dt$, is known as the Green-Kubo formula. It is obtained by expanding the definition of the variance of the time-average estimator, $\lim_{T \rightarrow \infty} T \cdot \text{Var}_\mu \left[\frac{1}{T} \int_0^T g(X_t)dt \right]$, and applying the stationarity of the process $(X_t)_{t \geq 0}$. The integrability of the auto-correlation function is guaranteed by the exponential decay of correlations, a consequence of the spectral gap of the process. For a detailed derivation, see, e.g., (Pavliotis, 2014, Chapter 4).

(b) The equivalence to the second formula: $\sigma_{g,J}^2 = 2 \langle g, (-\mathcal{L}_J)^{-1}g \rangle$, is established via the operator semigroup theory. The auto-correlation function can be expressed as $\mathbb{E}_\mu[g(X_0)g(X_t)] = \langle g, P_t^J g \rangle$, where $P_t^J = e^{t\mathcal{L}_J}$ is the Markov semigroup generated by \mathcal{L}_J . The Green-Kubo formula then becomes: $\sigma_{g,J}^2 = 2 \int_0^\infty \langle g, P_t^J g \rangle dt = 2 \langle g, (\int_0^\infty P_t^J dt) g \rangle$. The exchange of the inner product and the integral is justified by Fubini's theorem. The integral of the semigroup is the resolvent of the generator at zero, i.e., $\int_0^\infty P_t^J dt = (-\mathcal{L}_J)^{-1}$. This identity is fundamental in semigroup theory and connects the integrated time-evolution of the process to the inverse of its generator. The existence and boundedness of $(-\mathcal{L}_J)^{-1}$ on $L_0^2(\mu)$ is ensured by the spectral gap of \mathcal{L}_J on the compact domain K . This establishes the equivalence.

Appendix C. Derivation of Adjoint and Operator Decompositions under Assumption 3

In this appendix, we derive the explicit expressions for the adjoint operator \mathcal{L}_J^* and the symmetric and skew-symmetric parts, \mathcal{L}_S and \mathcal{L}_A , of the infinitesimal generator \mathcal{L}_J with respect to the invariant measure $\mu(dx) = Z^{-1}e^{-f(x)}dx$. These derivations are performed under Assumptions 1, 2, and crucially, Assumption 3. The infinitesimal generator of $(X_t)_{t \geq 0}$

in (2.1) is given by $\mathcal{L}_J u(x) = \Delta u(x) - \langle \nabla f(x), \nabla u(x) \rangle - \langle J(x) \nabla f(x), \nabla u(x) \rangle$, and acts on functions $u \in D(\mathcal{L}_J)$, which satisfy Neumann boundary conditions $\langle \nabla u(x), \mathbf{n}(x) \rangle = 0$ on ∂K due to Assumption 2 (where $\mathbf{n}(x)$ is the outward unit normal vector).

Let $\mathcal{O}_1 u := \Delta u - \langle \nabla f, \nabla u \rangle$ and $\mathcal{O}_2 u := -\langle J \nabla f, \nabla u \rangle$, so that $\mathcal{L}_J = \mathcal{O}_1 + \mathcal{O}_2$.

C.1 Adjoint of \mathcal{O}_1

The operator $\mathcal{O}_1 u = \Delta u - \langle \nabla f, \nabla u \rangle$ can be written as $\mathcal{O}_1 u = e^{f(x)} \nabla \cdot (e^{-f(x)} \nabla u(x))$. This is the infinitesimal generator of a reversible Langevin dynamics with invariant measure μ . For $u, v \in D(\mathcal{L}_J)$ (hence satisfying Neumann boundary conditions), \mathcal{O}_1 is self-adjoint with respect to μ :

$$\begin{aligned} \langle \mathcal{O}_1 u, v \rangle_\mu &= \frac{1}{Z} \int_K \left(e^f \nabla \cdot (e^{-f} \nabla u) \right) v e^{-f} dx = \frac{1}{Z} \int_K \left(\nabla \cdot (e^{-f} \nabla u) \right) v dx \\ &= \frac{1}{Z} \left[- \int_K (e^{-f} \nabla u) \cdot \nabla v dx + \int_{\partial K} v (e^{-f} \nabla u) \cdot \mathbf{n}_{out} dS \right] \\ &= -\frac{1}{Z} \int_K e^{-f} (\nabla u \cdot \nabla v) dx \quad (\text{since } \langle \nabla u, \mathbf{n}_{out} \rangle = 0). \end{aligned}$$

This expression is symmetric in u and v , so that $\langle \mathcal{O}_1 u, v \rangle_\mu = \langle u, \mathcal{O}_1 v \rangle_\mu$. Thus, $\mathcal{O}_1^* = \mathcal{O}_1$.

C.2 Adjoint of \mathcal{O}_2 under Assumption 3

The operator $\mathcal{O}_2 u = -\langle J \nabla f, \nabla u \rangle$. We seek \mathcal{O}_2^* such that $\langle \mathcal{O}_2 u, v \rangle_\mu = \langle u, \mathcal{O}_2^* v \rangle_\mu$.

$$\begin{aligned} \langle \mathcal{O}_2 u, v \rangle_\mu &= -\frac{1}{Z} \int_K \langle J(x) \nabla f(x), \nabla u(x) \rangle v(x) e^{-f(x)} dx \\ &= -\frac{1}{Z} \int_K \left(v(x) J(x) \nabla f(x) e^{-f(x)} \right) \cdot \nabla u(x) dx. \end{aligned}$$

Using integration by parts (divergence theorem, $\int_K \mathbf{F} \cdot \nabla u dx = -\int_K u(\nabla \cdot \mathbf{F}) dx + \int_{\partial K} u(\mathbf{F} \cdot \mathbf{n}_{out}) dS$, with $\mathbf{F}(x) = v(x) J(x) \nabla f(x) e^{-f(x)}$):

$$\begin{aligned} \langle \mathcal{O}_2 u, v \rangle_\mu &= -\frac{1}{Z} \left[- \int_K u(x) \nabla \cdot \left(v(x) J(x) \nabla f(x) e^{-f(x)} \right) dx \right. \\ &\quad \left. + \int_{\partial K} u(x) v(x) \left(J(x) \nabla f(x) e^{-f(x)} \right) \cdot \mathbf{n}_{out}(x) dS \right]. \end{aligned}$$

The boundary term is $\int_{\partial K} u v e^{-f} (J \nabla f \cdot \mathbf{n}_{out}) dS$. Since $\mathbf{n}_{out} = -\mathbf{n}_{in}$ and by Assumption 2 ($J \mathbf{n}_{in} = 0$), we have

$$\begin{aligned} J \nabla f \cdot \mathbf{n}_{out} &= - (J \nabla f \cdot \mathbf{n}_{in}) = - \left(\nabla f \cdot J^\top \mathbf{n}_{in} \right) = - (\nabla f \cdot (-J \mathbf{n}_{in})) \\ &= (\nabla f \cdot J \mathbf{n}_{in}) = (\nabla f \cdot \mathbf{0}) = 0. \end{aligned}$$

Thus, the boundary term vanishes. Therefore,

$$\langle \mathcal{O}_2 u, v \rangle_\mu = \frac{1}{Z} \int_K u(x) \nabla \cdot \left(v(x) J(x) \nabla f(x) e^{-f(x)} \right) dx.$$

The adjoint operator \mathcal{O}_2^* acting on a function v is thus:

$$\begin{aligned}\mathcal{O}_2^*v(x) &= e^{f(x)}\nabla \cdot \left(v(x)J(x)\nabla f(x)e^{-f(x)} \right) \\ &= e^{f(x)} \left[(\nabla v(x)) \cdot \left(J(x)\nabla f(x)e^{-f(x)} \right) + v(x)\nabla \cdot \left(J(x)\nabla f(x)e^{-f(x)} \right) \right] \\ &= \langle J(x)\nabla f(x), \nabla v(x) \rangle + v(x)e^{f(x)} \left(\nabla \cdot \left(J(x)\nabla f(x)e^{-f(x)} \right) \right).\end{aligned}$$

By chain rule and product rule, we can compute that

$$\begin{aligned}\nabla \cdot \left(J(x)\nabla f(x)e^{-f(x)} \right) &= -\nabla \cdot \left(J(x)\nabla \left(e^{-f(x)} \right) \right) = -\sum_i \frac{\partial}{\partial x_i} \left(\sum_k J_{ik}(x) \left(\frac{\partial}{\partial x_k} e^{-f(x)} \right) \right) \\ &= -\sum_i \left(\frac{\partial}{\partial x_i} \sum_k J_{ik}(x) \right) \left(\frac{\partial}{\partial x_k} e^{-f(x)} \right) - \sum_i \sum_k J_{ik}(x) \frac{\partial^2 e^{-f(x)}}{\partial x_i \partial x_k}.\end{aligned}$$

Under Assumption 3, we have $\nabla \cdot J(x) = 0$. Consequently,

$$\sum_i \left(\frac{\partial}{\partial x_i} \sum_k J_{ik}(x) \right) \left(\frac{\partial}{\partial x_k} e^{-f(x)} \right) = 0.$$

Moreover, $J(x)$ is skew-symmetric, and $\left(\frac{\partial^2 e^{-f(x)}}{\partial x_i \partial x_k} \right)_{i,k}$ is a symmetric matrix, which implies that $\sum_i \sum_k J_{ik}(x) \frac{\partial^2 e^{-f(x)}}{\partial x_i \partial x_k} = 0$. Hence, we conclude that $\nabla \cdot \left(J(x)\nabla f(x)e^{-f(x)} \right) = 0$, and the expression for $\mathcal{O}_2^*v(x)$ simplifies to:

$$\mathcal{O}_2^*v(x) = \langle J(x)\nabla f(x), \nabla v(x) \rangle.$$

C.3 Decomposition of \mathcal{L}_J under Assumption 3

Now we can compute \mathcal{L}_S and \mathcal{L}_A using the simplified \mathcal{O}_2^* .

The Symmetric Part \mathcal{L}_S :

$$\begin{aligned}\mathcal{L}_S u &= \frac{1}{2}(\mathcal{L}_J u + \mathcal{L}_J^* u) = \frac{1}{2}((\mathcal{O}_1 u + \mathcal{O}_2 u) + (\mathcal{O}_1^* u + \mathcal{O}_2^* u)) \\ &= \frac{1}{2}(\mathcal{O}_1 u + \mathcal{O}_2 u + \mathcal{O}_1 u + \mathcal{O}_2^* u) = \mathcal{O}_1 u + \frac{1}{2}(\mathcal{O}_2 u + \mathcal{O}_2^* u). \quad (\text{since } \mathcal{O}_1^* = \mathcal{O}_1)\end{aligned}$$

Substituting the expressions for $\mathcal{O}_1 u$, $\mathcal{O}_2 u = -\langle J\nabla f, \nabla u \rangle$, and the simplified $\mathcal{O}_2^* u = \langle J\nabla f, \nabla u \rangle$:

$$\mathcal{O}_2 u + \mathcal{O}_2^* u = -\langle J(x)\nabla f(x), \nabla u(x) \rangle + \langle J(x)\nabla f(x), \nabla u(x) \rangle = 0.$$

Thus, the symmetric part is:

$$\mathcal{L}_S u(x) = \mathcal{O}_1 u(x) = \Delta u(x) - \langle \nabla f(x), \nabla u(x) \rangle.$$

The Skew-Symmetric Part \mathcal{L}_A :

$$\mathcal{L}_A u = \frac{1}{2}(\mathcal{L}_J u - \mathcal{L}_J^* u) = \frac{1}{2}((\mathcal{O}_1 u + \mathcal{O}_2 u) - (\mathcal{O}_1 u + \mathcal{O}_2^* u)) = \frac{1}{2}(\mathcal{O}_2 u - \mathcal{O}_2^* u).$$

Substituting $\mathcal{O}_2 u = -\langle J \nabla f, \nabla u \rangle$ and the simplified $\mathcal{O}_2^* u = \langle J \nabla f, \nabla u \rangle$:

$$\mathcal{O}_2 u - \mathcal{O}_2^* u = -\langle J(x) \nabla f(x), \nabla u(x) \rangle - \langle J(x) \nabla f(x), \nabla u(x) \rangle = -2\langle J(x) \nabla f(x), \nabla u(x) \rangle.$$

Thus, the skew-symmetric part is: $\mathcal{L}_A u(x) = -\langle J(x) \nabla f(x), \nabla u(x) \rangle$. Under Assumption 3, \mathcal{L}_A is anti-self-adjoint with respect to μ .

References

- Kwangjun Ahn and Sinho Chewi. Efficient constrained sampling via the mirror-Langevin algorithm. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 28405–28418. Curran Associates, Inc., 2021.
- Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1):5–43, 2003.
- Raghu Raj Bahadur. Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics*, 38(2):303–324, 1967.
- Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.
- Nicolas Brosse, Alain Durmus, Éric Moulines, and Marcelo Pereyra. Sampling from a log-concave distribution with compact support with proximal Langevin Monte Carlo. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65, pages 319–342. PMLR, 2017.
- Sebastien Bubeck, Ronen Eldan, and Joseph Lehec. Finite-time analysis of projected Langevin Monte Carlo. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.
- Sébastien Bubeck, Ronen Eldan, and Joseph Lehec. Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete & Computational Geometry*, 59(4):757–783, 2018.
- Sinho Chewi, Thibaut Le Gouic, Cheng Lu, Tyler Maunu, Philippe Rigollet, and Austin Stromme. Exponential ergodicity of mirror-Langevin diffusions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 19573–19585. Curran Associates, Inc., 2020.
- E. Brian Davies. *Heat Kernels and Spectral Theory*, volume 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1989.
- Klaus Deimling. *Nonlinear Functional Analysis*. Springer Science & Business Media, 2013.
- Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 2nd edition, 1998.

- Monroe D. Donsker and S. R. Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, I. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975a.
- Monroe D. Donsker and S. R. Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, II. *Communications on Pure and Applied Mathematics*, 28(2):279–301, 1975b.
- Monroe D. Donsker and S. R. Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, III. *Communications on Pure and Applied Mathematics*, 29(4):389–461, 1976.
- Monroe D. Donsker and S. R. Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, IV. *Communications on Pure and Applied Mathematics*, 36(2):183–212, 1983.
- Hengrong Du, Qi Feng, Changwei Tu, Xiaoyu Wang, and Lingjiong Zhu. Non-reversible Langevin algorithms for constrained sampling. *arXiv:2501.11743*, 2025.
- Andrew B. Duncan, Tony Lelièvre, and Grigoris A. Pavliotis. Variance reduction using nonreversible Langevin samplers. *Journal of Statistical Physics*, 163(3):457–491, 2016.
- Andrew B. Duncan, Nikolas Nüsken, and Grigoris A. Pavliotis. Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions. *Journal of Statistical Physics*, 169(6):1098–1131, Dec 2017.
- Ivar Ekeland and Roger Temam. *Convex Analysis and Variational Problems*. SIAM, 1999.
- Lawrence C Evans. *Partial Differential Equations*, volume 19. American Mathematical Society, 2022.
- Grégoire Ferré and Gabriel Stoltz. Large deviations of empirical measures of diffusions in weighted topologies. *Electronic Journal of Probability*, 25(121):1–52, 2020.
- Xuefeng Gao, Mert Gürbüzbalaban, and Lingjiong Zhu. Breaking reversibility accelerates Langevin dynamics for global non-convex optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 17850–17862. Curran Associates, Inc., 2020.
- Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC Press, 1995.
- David Gilbarg and Neil S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer, Berlin and Heidelberg, 2001.
- Arnaud Guillin and Pierre Monmarché. Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electronic Communications in Probability*, 21:14 pp., 2016.
- Mert Gürbüzbalaban, Xuefeng Gao, Yuanhan Hu, and Lingjiong Zhu. Decentralized stochastic gradient Langevin dynamics and Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 22(239):1–69, 2021.

- Mert Gürbüzbalaban, Yuanhan Hu, and Lingjiong Zhu. Penalized overdamped and underdamped Langevin Monte Carlo algorithms for constrained sampling. *Journal of Machine Learning Research*, 25(263):1–67, 2024a.
- Mert Gürbüzbalaban, Mohammad Rafiqul Islam, Xiaoyu Wang, and Lingjiong Zhu. Generalized EXTRA stochastic gradient Langevin dynamics. *arXiv preprint arXiv:2412.01993*, 2024b.
- Ya-Ping Hsieh, Ali Kavis, Paul Rolland, and Volkan Cevher. Mirrored Langevin dynamics. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- Yuanhan Hu, Xiaoyu Wang, Xuefeng Gao, Gürbüzbalaban, and Lingjiong Zhu. Non-convex stochastic optimization via non-reversible stochastic gradient Langevin dynamics. *arXiv:2004.02823*, 2020.
- Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. Accelerating Gaussian diffusions. *Annals of Applied Probability*, 3:897–913, 1993.
- Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. Accelerating diffusions. *Annals of Applied Probability*, 15:1433–1444, 2005.
- Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, 2013.
- Yunbum Kook, Yin-Tat Lee, Ruoqi Shen, and Santosh Vempala. Sampling with Riemannian Hamiltonian Monte Carlo in a constrained space. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 31684–31696. Curran Associates, Inc., 2022.
- Andrew Lamperski. Projected stochastic gradient Langevin algorithms for constrained sampling and non-convex learning. In *Conference on Learning Theory*, volume 134, pages 2891–2937. PMLR, 2021.
- Dorian Le Peutrec and Laurent Michel. Sharp spectral asymptotics for nonreversible metastable diffusion processes. *Probability and Mathematical Physics*, 1(1):3–53, 2020.
- Jungkyoung Lee and Insuk Seo. Non-reversible metastable diffusions with Gibbs invariant measure I: Eyring–Kramers formula. *Probability Theory and Related Fields*, 182(3):849–903, 2022.
- Tony Lelièvre, Francis Nier, and Grigorios A Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.
- Ruilin Li, Molei Tao, Santosh S. Vempala, and Andre Wibisono. The mirror Langevin algorithm converges with vanishing bias. In *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, volume 167, pages 718–742. PMLR, 2022.

- Gary M Lieberman. Local estimates for subsolutions and supersolutions of oblique derivative problems for general second order elliptic equations. *Transactions of the American Mathematical Society*, 304(1):343–353, 1987.
- Grigorios A Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, volume 60. Springer, 2014.
- Amnon Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44. Springer Science & Business Media, 1983.
- Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviation approach. *Nonlinearity*, 28:2081–2103, 2015.
- Adil Salim and Peter Richtárik. Primal dual interpretation of the proximal stochastic gradient Langevin algorithm. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 3786–3796. Curran Associates, Inc., 2020.
- Mark Schmidt. Least squares optimization with L1-norm regularization. *CS542B Project Report*, 504(2005):195–221, 2005.
- Daniel W Stroock. *Partial Differential Equations for Probabalists*, volume 112. Cambridge University Press, 2008.
- Andrew M Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.
- Yee Whye Teh, Alexandre H Thiery, and Sebastian J Vollmer. Consistency and fluctuations for stochastic gradient Langevin dynamics. *Journal of Machine Learning Research*, 17(1):193–225, 2016.
- S. R. Srinivasa Varadhan. *Large Deviations and Applications*. SIAM, Philadelphia, 1984.
- Sheng-Jhih Wu, Chii-Ruey Hwang, and Moody T. Chu. Attaining the optimal Gaussian diffusion acceleration. *Journal of Statistical Physics*, 155(3):571–590, 2014.
- Nian Yao, Pervez Ali, Xihua Tao, and Lingjiong Zhu. Accelerating Langevin Monte Carlo sampling: A large deviations analysis. *arXiv:2503.19066*, 2025.
- Kelvin Shuangjian Zhang, Gabriel Peyré, Jalal Fadili, and Marcelo Pereyra. Wasserstein control of mirror Langevin Monte Carlo. In *Conference on Learning Theory*, volume 125, pages 3814–3841. PMLR, 2020.
- Haoyang Zheng, Hengrong Du, Qi Feng, Wei Deng, and Guang Lin. Constrained exploration via reflected replica exchange stochastic gradient Langevin dynamics. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 61321–61348. PMLR, 2024.
- Yuping Zheng and Andrew Lamperski. Constrained Langevin algorithms with L-mixing external random variables. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 20511–20521. Curran Associates, Inc., 2022.