

Variance Reduction Using Nonreversible Langevin Samplers

A. B. Duncan¹  · T. Lelièvre² · G. A. Pavliotis¹

Received: 16 June 2015 / Accepted: 29 February 2016 / Published online: 22 March 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract A standard approach to computing expectations with respect to a given target measure is to introduce an overdamped Langevin equation which is reversible with respect to the target distribution, and to approximate the expectation by a time-averaging estimator. As has been noted in recent papers [30,37,61,72], introducing an appropriately chosen nonreversible component to the dynamics is beneficial, both in terms of reducing the asymptotic variance and of speeding up convergence to the target distribution. In this paper we present a detailed study of the dependence of the asymptotic variance on the deviation from reversibility. Our theoretical findings are supported by numerical simulations.

1 Introduction

1.1 Motivation

In various applications arising in statistical mechanics, biochemistry, data science and machine learning [19,38,39,42], it is often necessary to compute expectations

$$\pi(f) := \mathbb{E}_\pi f = \int_{\mathbb{R}^d} f(x)\pi(dx)$$

✉ A. B. Duncan
a.duncan@imperial.ac.uk

T. Lelièvre
lelievre@cermics.enpc.fr

G. A. Pavliotis
g.pavliotis@imperial.ac.uk

¹ Department of Mathematics, South Kensington Campus, Imperial College London, London SW7 2AZ, England

² CERMICS, Ecole des ponts, Université Paris-Est, 6-8 avenue Blaise Pascal, 77455 Marne la Vallée Cedex 2, France

of an observable f with respect to a target probability distribution $\pi(dx)$ on \mathbb{R}^d with density $\pi(x)$ with respect to the Lebesgue measure, known up to the normalization constant.¹ When the dimension d is large, standard deterministic quadrature approaches become intractable, and one typically resorts to Markov-chain Monte Carlo (MCMC) methods [19,39,63]. In this approach, $\pi(f)$ is approximated by a long-time average of the form:

$$\pi_T(f) := \frac{1}{T} \int_0^T f(X_t) dt, \tag{1}$$

where X_t is a Markov process chosen to be ergodic with respect to the target distribution π . The Birkhoff–von Neumann Ergodic theorem [31,34,60] states that, for every observable $f \in L^1(\pi)$ we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(X_s) ds = \pi(f), \quad \pi - \text{ a.e. } X_0 = x. \tag{2}$$

If π possesses a smooth, strictly positive density, then a natural choice for X_t is the overdamped Langevin dynamics

$$dX_t = \nabla \log \pi(X_t) dt + \sqrt{2} dW_t, \tag{3}$$

where W_t is a standard Brownian motion on \mathbb{R}^d . Assuming that (3) possesses a unique strong solution which is non-explosive, the process X_t is ergodic, with unique invariant distribution π , such that (2) holds. Under additional assumptions on the distribution π and on the observable, this convergence result is accompanied by a central limit theorem which characterizes the asymptotic distribution of the fluctuations of $\pi_T(f)$ about $\pi(f)$, i.e.

$$\sqrt{T} (\pi_T(f) - \pi(f)) \xrightarrow{D} \mathcal{N}(0, \sigma_f^2), \tag{4}$$

where σ_f^2 is known as the asymptotic variance for the observable f . For the reversible process (3) started from stationarity (i.e. $X_0 \sim \pi$), the Kipnis–Varadhan theorem [12,32] implies that (4) holds with asymptotic variance

$$\sigma_f^2 = 2 \langle f - \mathbb{E}_\pi f, (-S)^{-1} (f - \mathbb{E}_\pi f) \rangle_\pi,$$

where $\langle \cdot, \cdot \rangle_\pi$ is the inner product in $L^2(\pi)$ and S is the infinitesimal generator of X_t defined by

$$S = \nabla \log \pi(x) \nabla \cdot + \Delta. \tag{5}$$

Sampling methods based on Langevin diffusions have become increasingly popular due to their wide applicability and relative ease of implementation. In practice, a discretisation of (3) may be used, which in general, will not be ergodic with respect to the target distribution π . Thus, the discretisation is augmented with a Metropolis–Hastings accept/reject step [65] which guarantees that the resulting Markov chain remains reversible with respect to π . The resulting scheme is known as the Metropolis-Adjusted Langevin Algorithm (MALA), see [65].

The computational cost required to approximate $\pi(f)$ using $\pi_T(f)$ to a given level of accuracy depends on the target distribution π , the observable f and the process X_t over which the long-time average is generated. Quite often, X_t will exhibit some form of metastability [10, 11, 36, 66]: the process X_t will remain trapped for a long time exploring one mode, with transitions between different modes occurring over longer timescales. When the observable

¹ With a slight abuse of notation, we will denote by π both the measure and the density.

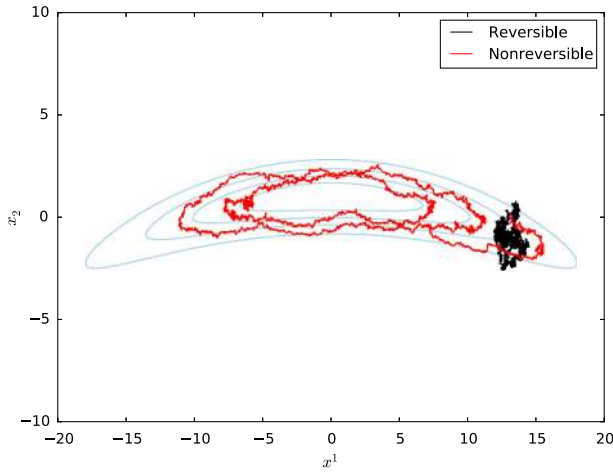


Fig. 1 Trajectories of a reversible MALA chain and a nonreversible Langevin sampler generating samples from a warped Gaussian distribution. The mass of the distribution is concentrated along a one-dimensional submanifold. Reversible samplers, such as MALA, perform a very slow exploration of the distribution, spending large amounts of time concentrated around a small region of the submanifold. On the other hand, nonreversible samplers are able to perform long “jumps” along the level-curves of the distribution, thus able to explore the distribution far more effectively (Color figure online)

depends directly on the metastable degrees of freedom (i.e. the observable takes different values in different metastable regions), the asymptotic variance σ_f^2 of the estimator $\pi_T(f)$ may be very large. As a result, more samples are required to obtain an estimate of $\pi(f)$ with the desired accuracy. A similar scenario arises when the mass of π is tightly concentrated along a low-dimensional submanifold of \mathbb{R}^d , as illustrated in Fig. 1. In this case, reversible dynamics, such as (3) will cause a very slow exploration of the support of π . As a result, $\pi_T(f)$ will exhibit very large variance for observables which vary strongly along the manifold.

As there are infinitely² many Markov processes with invariant distribution π , a natural question is whether such a process can be chosen to have optimal performance. The two standard optimality criteria that are commonly used are:

- (a) With respect to speeding up convergence to the target distribution.
- (b) With respect to minimizing the asymptotic variance.

These criteria can be used in order to introduce a partial ordering in the set of Markov chains or diffusion processes that are ergodic with respect to a given probability distribution [50, 59]. From a practical perspective, the definite optimality criterion is that of minimizing the computational cost. We address this issue (at least partially) later in this paper.

Within the family of reversible samplers, much work has been done to derive samplers which exhibit good (if not optimal) computational performance. This has motivated a number of variants of MALA which exploit geometric features of π to explore the state space more effectively, including preconditioned MALA [64], Riemannian Manifold MALA [20] and Stochastic Newton Methods [45].

² Formally, all diffusion processes X_t with drift $b(x)$ and diffusion $\sigma(x)$ and generator $\mathcal{L} = b \cdot \nabla + \frac{1}{2} \sigma \sigma^\top : \nabla \nabla$ such that π is the unique solution of the stationary Fokker-Planck equation $\mathcal{L}^* \pi = 0$ can be used to sample from π .

1.2 Nonreversible Langevin Dynamics

An MCMC scheme which departs from the assumption of reversible dynamics is Hamiltonian MCMC [53], which has proved successful in Bayesian inference. By augmenting the state space with a momentum variable, proposals for the next step of the Markov chain are generated by following Hamiltonian dynamics over a large, fixed time interval. The resulting nonreversible chain is able to make distant proposals. Various methods have been proposed which are related to this general idea of breaking nonreversibility by introducing an additional dimension to the state space and introducing dynamics which explore the enlarged space while still preserving the equilibrium distribution. In particular, the lifting method [15, 27, 71] is one such method applied to discrete state systems, where the Markov chain is “lifted” from the state space Ω onto the space $\Omega \times \{1, -1\}$. The transition probabilities in each copy are modified to introduce transitions between the copies to preserve the invariant distribution but now promote the sampler to generate long strides or trajectories. Similar methods based on introducing nonreversibility into the dynamics of discrete state chains to speed up convergence have been applied with success in various applications [9, 26, 52, 68, 69]. These methods are also reminiscent of parallel tempering or replica exchange MCMC [51], which are aimed at efficiently sampling from multimodal distributions.

It is well documented that breaking detailed balance, i.e. considering a nonreversible diffusion process that is ergodic with respect to π , can help accelerate convergence to equilibrium. In fact, it has been proved that, among all diffusion processes with additive noise that are ergodic with respect to π , the reversible dynamics (3) has the slowest rate of convergence to equilibrium, measured in terms of the spectral gap of the generator in $L^2(\mathbb{R}^d; \pi) =: L^2(\pi)$, c.f. [37]. Adding a drift to (3) that is divergence-free with respect to π and that preserves the invariant measure of the dynamics will always accelerate convergence to equilibrium [28, 29, 37, 54, 61, 72]. The optimal nonreversible perturbation can be identified and obtained in an algorithmic manner for diffusions with linear drift, whose invariant distribution is Gaussian [37]; see also [72].

The effect of nonreversibility in the dynamics on the asymptotic variance has also been studied. In [61] it is shown that small antisymmetric perturbations of the reversible dynamics always decrease the asymptotic variance, and more recently in [62] Friedlin–Wentzell theory is used to study the limit of infinitely strong antisymmetric perturbations. In [30] the authors use spectral theory for selfadjoint operators to study the effect of antisymmetric perturbations on diffusions on both \mathbb{R}^d and on compact manifolds and provide a general comparison result between reversible and nonreversible diffusions. This work is related with previous studies on the behavior of the spectral gap of the generator when the strength of the nonreversible perturbation is increased [13, 18].

1.3 Objectives of This Paper

In this paper we present a detailed analytical and computational study of the effect of adding a nonreversible drift to the reversible overdamped Langevin dynamics (3) on the asymptotic variance σ_f^2 . We will consider the nonreversible dynamics defined by [28, 29] :

$$dX_t^\gamma = (\nabla \log \pi(X_t^\gamma) + \alpha \gamma(X_t^\gamma)) dt + \sqrt{2} dW_t, \tag{6}$$

where the smooth vector field γ is taken to be divergence-free with respect to the distribution π ,

$$\nabla \cdot (\gamma \pi) = 0. \tag{7}$$

There are infinitely many vector fields that satisfy (7) and a general formula can be derived using Poincaré’s lemma [29, 62]. We can, for example, construct such vector fields by taking

$$\gamma(x) = J \nabla \log \pi(x), \quad J = -J^T. \tag{8}$$

In (6) we have already normalized the various vector fields and we have introduced a parameter α which measures the strength of the deviation from reversible dynamics. The generator of the diffusion process X_t^γ can be decomposed into a symmetric and an antisymmetric part in $L^2(\pi)$, representing the reversible and irreversible parts of the dynamics, respectively [56, Sect. 4.6]:

$$\mathcal{L} = \mathcal{S} + \alpha \mathcal{A}, \tag{9}$$

where \mathcal{S} is given in (5) and $\mathcal{A} = \gamma \cdot \nabla$. We are particularly interested in quantifying the reduction in the asymptotic variance obtained caused by breaking detailed balance. It is well known [32, 33] that for square integrable observables, the asymptotic variance, denoted by $\sigma_f^2(\alpha)$ can be written in terms of the solution of the Poisson equation

$$-(\mathcal{S} + \alpha \mathcal{A})\phi = f - \mathbb{E}_\pi f, \tag{10}$$

as

$$\sigma_f^2(\alpha) = 2 \int \phi(-\mathcal{L})\phi \pi(dx). \tag{11}$$

Our first goal is to obtain an explicit formula for $\sigma_f^2(\alpha)$ in terms of \mathcal{S} and \mathcal{A} . Moreover, through a spectral decomposition of the operator $(-\mathcal{S})^{-1}\mathcal{A}$ we investigate the dependence of the asymptotic variance on the strength of the nonreversible perturbation. Based on earlier work by Golden and Papanicolaou [22], Majda et al. [3, 44] and Bhattacharya et al. [7, 8], in [58], an expression is obtained for the effective self-diffusion coefficient for a tagged particle whose microscopic behaviour is determined by a nonreversible Markov process. The basic idea is the introduction of the operator

$$\mathcal{G} := (-\mathcal{S})^{-1}\mathcal{A}, \tag{12}$$

which is skewadjoint in the weighted Sobolev space

$$\mathcal{H}^1 = \{f \in L^2(\pi) : \pi(f) = 0 \text{ and } \langle (-\mathcal{S})f, f \rangle_\pi < \infty\},$$

where $\langle \cdot, \cdot \rangle_\pi$ denotes the $L^2(\pi)$ inner product; see Sect. 3. We follow a similar approach in this paper, from which a quite explicit formula for the asymptotic variance $\sigma_f^2(\alpha)$ is derived using the spectral theorem for selfadjoint operators. From this expression we can immediately deduce that the addition of a nonreversible perturbation reduces the asymptotic variance and we can also study the small and large α asymptotics of σ_f^2 ; see Theorem 4. One of the results that we prove is that the large α behaviour of the asymptotic variance depends on the detailed properties of the vector field γ : when the nullspace of the antisymmetric part of the generator $\mathcal{A} = \gamma \cdot \nabla$ consists of only constants in \mathcal{H}^1 , then the asymptotic variance converges to 0. Indeed, in this case, in the $|\alpha| \rightarrow \infty$ limit, the limiting dynamics become deterministic, characterized by the Liouville operator $\gamma \cdot \nabla$. On the other hand, when the nullspace of \mathcal{A} contains nontrivial functions in \mathcal{H}^1 there will exist observables for which the asymptotic variance σ_f^2 converges to a positive constant as $|\alpha| \rightarrow \infty$.

The effect of the antisymmetric part on the long time dynamics of diffusion processes has been also studied extensively in the context of turbulent diffusion [43] and fluid mixing. The effect of an incompressible flow on the convergence of the solution to the advection–diffusion equation on a compact manifold to its mean value (i.e. when $\pi \equiv 1$) was first studied in [13].

In particular, the concept of a relaxation enhancing flow was introduced and it was shown that a divergence-free flow is relaxation enhancing if and only if the Liouville operator $\mathcal{A} = v \cdot \nabla$ has no eigenfunctions in the Sobolev space H^1 . An equivalent formulation of this result is that an incompressible flow is relaxation enhancing if and only if it is weakly mixing. Examples of relaxation enhancing flows are given in [13]. This problem was studied further in [18], where it also mentioned that there are very few examples of relaxation enhancing flows. In [18] it is shown that the spectral gap of the advection-diffusion operator $\mathcal{L} = \alpha v \cdot \nabla + \Delta$, for a divergence-free drift v and $\alpha \in \mathbb{R}$, remains bounded above in the limit as $\alpha \rightarrow \pm\infty$ by a negative constant if and only if the advection operator has an eigenfunction in H^1 . These results are reminiscent of the results we mention above on the necessary and sufficient condition to obtain a reduction of asymptotic variance in the limit $\alpha \rightarrow \pm\infty$.

Our analysis of the asymptotic variance σ_f^2 , based on the careful study of the Poisson Eq. (10), enables us to study in detail the problem of finding the nonreversible perturbation giving rise to minimum asymptotic variance for diffusions with linear drift, i.e. diffusions whose invariant measure is Gaussian, over a large class of observables. Diffusions with linear drift were considered in [37] where the optimal nonreversible perturbation with respect to accelerating convergence was obtained. For linear and quadratic observables, we can give a complete solution to this problem, and construct nonreversible perturbations that provide a dramatic reduction in asymptotic variance. Moreover, we demonstrate that the conditions under which the variance is reduced are very different from those of maximising the spectral gap discussed in [37]. In particular, we show how a nonreversible perturbation can dramatically reduce the asymptotic variance for the estimator $\pi_T(f)$, even though no such improvement can be made on the rate of convergence to equilibrium.

Guided by our theoretical results, we can then study numerically the reduction in the asymptotic variance due to the addition of a nonreversible drift for some toy models from molecular dynamics. In particular, we study the problem of computing expectations of observables with respect to a warped Gaussian [23] in two dimensions, as well as a simple model for a dimer in a solvent [38]. The numerical experiments reported in this paper illustrate that a judicious choice of the nonreversible perturbation, dependent on the target distribution and the observable, can dramatically reduce the asymptotic variance.

To compute $\pi_T(f)$ numerically, we use an Euler–Maruyama discretisation of (6). The resulting discretisation error introduces an additional bias in the estimator for $\pi(f)$, see [47] for a comprehensive error analysis. This imposes additional constraints on the magnitude of the nonreversible drift, since increasing α arbitrarily will give rise to a large discretisation error which must be controlled by taking smaller timesteps. A natural question is whether the increase in the computational cost due to the necessity of taking smaller timesteps negates any benefits of the resulting variance reduction. To study this problem, we compare the computational cost of the unadjusted nonreversible Langevin sampler with the corresponding MALA scheme.³ Our numerical results, together with the theoretical analysis for diffusions with linear drift, show that the nonreversible Langevin sampler can outperform the MALA algorithm, provided that the nonreversible perturbation is well-chosen. Finally, we consider a higher order numerical scheme for generating samples of (6), based on splitting the reversible and nonreversible dynamics. Numerically, we investigate the properties of this integrator, and demonstrate that its improved stability and discretisation error make it a good numerical scheme for computing the estimator $\pi_T(f)$ using a nonreversible diffusion.

³ As we illustrate in our paper, there is no point in considering the Metropolis adjusted sampler with a non-reversible proposal, since the addition of the accept–reject step renders the resulting Markov chain reversible and any nonreversibility-induced variance reduction is lost.

The rest of the paper is organized as follows. In Sect. 2 we describe how the central limit theorem (4) arises from the solution of the Poisson equation associated with the generator of the dynamics. In Sect. 3 we analyse the asymptotic variance and formulate the problem of finding the optimal perturbation with respect to minimising σ_f^2 , for a fixed observable, or over the space of square-integrable observables. Moreover, following the analysis described in [58], we derive a spectral representation for the asymptotic variance σ_f^2 , in terms of the discrete spectrum of the operator \mathcal{G} defined in (12), and recover estimates for the asymptotic variance for any value of α . In Sect. 4, we consider the case of Gaussian diffusions, which are amenable to explicit calculation to demonstrate the theory presented in this paper. In Sect. 5, we provide various numerical examples to complement the theoretical results. Finally, in Sect. 6 we describe the bias-variance tradeoff for nonreversible Langevin samplers, and explore their computational cost. Conclusions and discussion on further work are presented in Sect. 7.

2 The Central Limit Theorem and Estimates on the Asymptotic Variance via the Poisson Equation

In this section we make explicit some sufficient conditions under which the estimator

$$\pi_T(f) = \frac{1}{T} \int_0^T f(X_s^\gamma) ds,$$

where X_t^γ denotes the solution of (6), satisfies a central limit theorem of the form (4). The fundamental ingredient for proving such a central limit theorem is establishing the well-posedness of the Poisson equation

$$-\mathcal{L}\phi(x) = f(x) - \pi(f), \quad \pi(\phi) = 0, \tag{13}$$

for all bounded and continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$, where \mathcal{L} is defined by (9), and obtaining estimates on the growth of the unique solution ϕ . As described previously, we shall assume that π possesses a smooth, strictly positive density, also denoted by $\pi(x)$, such that $\int_{\mathbb{R}^d} \pi(x) dx < \infty$ and that the SDE (6) has a unique strong solution for all $\alpha \in \mathbb{R}$. Applying the results detailed in [21,49], we shall assume that the process X_t^γ possesses a Lyapunov function, which is sufficient to ensure the exponential ergodicity of X_t^γ , [46,48], as detailed in the following assumption.

Assumption A (Foster–Lyapunov Criterion) There exists a function $U : \mathbb{R}^d \rightarrow \mathbb{R}$ and constants $c > 0$ and $b \in \mathbb{R}$ such that $\pi(U) < \infty$ and

$$\mathcal{L}U(x) \leq -cU(x) + b\mathbf{1}_C, \quad \text{and} \quad U(x) \geq 1, \quad x \in \mathbb{R}^d, \tag{14}$$

where $\mathbf{1}_C$ is the indicator function over a *petite set*. For the definition of a petite set we refer the reader to [48]. For the generator \mathcal{L} corresponding to the process (6), compact sets are always petite. As noted in [65], a sufficient condition on π for (6) to possess a Lyapunov function is the following.

Assumption B The density π is bounded and for some $0 < \beta < 1$:

$$\liminf_{|x| \rightarrow \infty} ((1 - \beta)|\nabla \log \pi(x)|^2 + \Delta \log \pi(x)) > 0. \tag{15}$$

Lemma 1 [65, Theorem 3.3] *Suppose that Assumption B holds then the Foster–Lyapunov criterion holds for (3) with*

$$U(x) = \pi^{-\beta}(x).$$

Moreover, if the nonreversible term is of the form $\gamma(x) = J \nabla \log \pi(x)$, for $J \in \mathbb{R}^{d \times d}$ antisymmetric, then this choice of U is a Lyapunov function for X_t^γ defined by (6) for all $\alpha \in \mathbb{R}$.

Proof For this choice of γ the generator of (6) has the form

$$\mathcal{L} = (I + \alpha J) \nabla \log \pi(x) \cdot \nabla + \Delta, \quad \alpha \in \mathbb{R},$$

For $U(x) = \pi^{-\beta}(x)$ we obtain:

$$\begin{aligned} \mathcal{L}U(x) &= -\beta \pi^{-\beta}(x) \nabla \log \pi(x) \cdot (I + \alpha J) \nabla \log \pi(x) - \beta \nabla \cdot (\pi^{-\beta}(x) \nabla \log \pi(x)) \\ &= -\beta [(1 - \beta) |\nabla \log \pi(x)|^2 + \Delta \log \pi(x)] \pi^{-\beta}(x). \end{aligned}$$

Thus, by assumption (15), there exists $\epsilon > 0$ and M such that for $|x| > M$:

$$(1 - \beta) |\nabla \log \pi(x)|^2 + \Delta \log \pi(x) > \epsilon,$$

and so

$$\mathcal{L}U(x) \leq -\beta \epsilon U(x) + b \mathbf{1}_{C_M},$$

where $C_M = \{x \in \mathbb{R}^d : |x| \leq M\}$ and b is a positive constant.

Finally, we note that since π is bounded, then U is bounded above from zero uniformly. Thus, U can be rescaled to satisfy the condition $U \geq 1$, as is required by the Foster–Lyapunov criterion. □

Remark 1 Note that Assumption B holds trivially when π is a Gaussian distribution in \mathbb{R}^d . More generally, following [64, Example 1], consider $\pi(x) \propto \exp(-p(x))$, where $p(x)$ is a polynomial of order m such that $p(x) \rightarrow \infty$ as $|x| \rightarrow \infty$ (necessarily $m \geq 2$ and m is even). Clearly

$$\liminf_{|x| \rightarrow \infty} \frac{|\nabla \log \pi(x)|^2}{|\Delta \log \pi(x)|} = \infty,$$

and

$$\liminf_{|x| \rightarrow \infty} (1 - \beta) |\nabla \log \pi(x)|^2 > 0,$$

so that (14) holds. On the other hand, consider the case when π is a Student’s t -distribution, $\pi(x) \propto (1 + x^2/\nu)^{-\frac{\nu+1}{2}}$ with $\nu \geq 2$. Then it is straightforward to check that (14) will not hold. Indeed, since $|\nabla \log \pi(x)| \rightarrow 0$, by [65, Theorem 2.4] this process is not exponentially ergodic.

If condition (14) holds for X_t^γ , then the process will be exponentially ergodic. More specifically, the law of the process X_t^γ started from a point $x \in \mathbb{R}^d$ will converge exponentially fast in the total variation norm to the equilibrium distribution π . In particular, denoting by $(P_t^\gamma)_{t \geq 0}$ the semigroup associated with the diffusion process (6), we have the following result.

Theorem 1 [12, Thm 8.3], [16, Thm 3.10, 3.12], [17, Thm 5.2.c] *Suppose that Assumption A holds for (6), with Lyapunov function U . Then there exist positive constants c and λ such that, for all x ,*

$$\|P_t^\gamma(x, \cdot) - \pi\|_{TV} \leq cU(x)e^{-\lambda t},$$

where $p_t^\gamma(x, \cdot)$ denotes the law of X_t^γ given $X_0^\gamma = x$ and $\|\cdot\|_{TV}$ denotes the total variation norm. In particular, for any probability measure ν such that $U \in L^1(\nu)$,

$$\lim_{t \rightarrow +\infty} \|(P_t^\gamma)^* \nu - \pi\|_{TV} = 0,$$

where $(P_t^\gamma)^*$ denotes the $L^2(\mathbb{R}^d)$ adjoint of P_t^γ . □

For a central limit theorem to hold for the process X_t^γ , and thus for σ_f^2 to be finite, it is necessary that the Poisson Eq. (10) is well-posed. The Foster-Lyapunov condition (14) is sufficient for this to hold.

Theorem 2 [21, Thm3.2] *Suppose that Assumption A holds for the diffusion process (6) with Lyapunov function U . Then there exists a positive constant c such that for any $|f|^2 \leq U$, the Poisson Eq. (10) admits a unique zero mean solution ϕ satisfying the bound $|\phi(x)|^2 \leq cU(x)$. In particular, $\phi \in L^2(\pi)$.* □

The technique of using a Poisson equation to obtain a central limit theorem for an additive functional of a Markov process is widely known, [5, 6, 55]. The approach is based on the fact that, at least formally, we can decompose $\pi_t(f) - \pi(f)$ into a martingale and a “remainder” term:

$$\begin{aligned} \pi_t(f) - \pi(f) &= \frac{1}{t} \int_0^t f(X_s^\gamma) ds - \pi(f) = \frac{\phi(X_0^\gamma) - \phi(X_t^\gamma)}{t} + \frac{\sqrt{2}}{t} \int_0^t \nabla \phi(X_s^\gamma) \cdot dW_s \\ &=: R_t + M_t. \end{aligned}$$

Considering the rescaling $\sqrt{t}(\pi_t(f) - \pi(f))$, the martingale term $\sqrt{t}M_t$ will converge in distribution to a Gaussian random variable with mean 0 and variance

$$\sigma_f^2 = 2 \int_{\mathbb{R}^d} |\nabla \phi(x)|^2 \pi(dx),$$

by the central limit theorem for martingales [25]. It remains to control the remainder term $\sqrt{t}R_t$. We distinguish between two cases: If $X_0^\gamma \sim \pi$, then since $\phi \in L^2(\pi)$, $\sqrt{t}R_t$ converges to 0 in $L^2(\pi)$ and the result follows. In the more general case we must resort to a “propagation of chaos” argument (c.f. [12, Sect. 8]), i.e. apply Theorem 1 to show that

$$\begin{aligned} &\mathbb{E}_{X_0=x} \left[H \left(\frac{1}{\sqrt{t}} \int_r^{t+r} f(X_s^\gamma) - \pi(f) ds \right) \right] \\ &- \mathbb{E}_{X_0 \sim \pi} \left[H \left(\frac{1}{\sqrt{t}} \int_0^t f(X_s^\gamma) - \pi(f) ds \right) \right] \rightarrow 0, \end{aligned}$$

as $r \rightarrow \infty$, for all continuous bounded functions H . The result then follows by decomposing

$$\pi_{t+r}(f) - \pi(f) = \frac{1}{t} \int_0^r f(X_s^\gamma) ds + \frac{1}{t} \int_r^{t+r} f(X_s^\gamma) ds,$$

and applying the propagation of chaos argument to the second term. The conclusion is summarized in the following result, which provides a central limit theorem for X_t^γ starting from an arbitrary initial distribution ν .

Theorem 3 [21, Thm 4.4] *If Assumption A holds for Lyapunov function U , then for any f such that $f^2(x) \leq U(x)$, there exists a constant $0 < \sigma_f^2 < \infty$ such that $\sqrt{t}(\pi_t(f) - \pi(f))$*

converges in distribution to an $\mathcal{N}(0, \sigma_f^2)$ distribution, as $t \rightarrow \infty$, for any initial distribution ν , where

$$\sigma_f^2 = 2 \int_{\mathbb{R}^d} |\nabla \phi(x)|^2 \pi(dx). \tag{16}$$

□

In the remainder of this paper we shall study the dependence of ϕ , and thus σ_f^2 on the choice of non-reversible perturbation γ . We note that (16) is precisely the Dirichlet form associated with the dynamics \mathcal{L} evaluated at the solution ϕ of the Poisson Eq. (13).

3 Analysis of the Asymptotic Variance

3.1 Mathematical Setting

We present a few definitions from [33] that will be useful in the sequel. Let $L_0^2(\pi)$ denote the set of $L^2(\pi)$ -integrable functions with zero mean, with corresponding inner product $\langle \cdot, \cdot \rangle_\pi$ and norm $\|\cdot\|_\pi$. Consider the operator \mathcal{S} given by (5) densely defined on $L_0^2(\pi)$. For $k \in \mathbb{N}$, given the family of seminorms

$$\|f\|_k^2 := \langle f, (-\mathcal{S})^k f \rangle_\pi, \tag{17}$$

we define the function spaces

$$\mathcal{H}^k := \{f \in L_0^2(\pi) : \|f\|_k < +\infty\}.$$

It follows that $\|\cdot\|_k$ is a norm on \mathcal{H}^k . For $k \geq 1$, the norm $\|\cdot\|_k$ satisfies the parallelogram identity and, consequently, the completion of \mathcal{H}^k with respect to the norm $\|\cdot\|_k$, which is denoted by \mathcal{H}^k , is a Hilbert space. The inner product $\langle \cdot, \cdot \rangle_k$ in \mathcal{H}^k is defined through polarization. It is easy to check that, for $f, g \in \mathcal{D}(\mathcal{L})$,

$$\langle f, g \rangle_1 = \langle f, (-\mathcal{S})g \rangle_\pi = \langle \nabla f, \nabla g \rangle_\pi. \tag{18}$$

We note that $\mathcal{H}^0 = L_0^2(\pi)$. A careful analysis of the function space \mathcal{H}^k is presented in [33].

The operator \mathcal{S} is symmetric with respect to π and can be extended to a selfadjoint operator on $L_0^2(\pi)$, which is also denoted by \mathcal{S} , with domain $\mathcal{D}(\mathcal{S}) = \mathcal{H}^2$. We shall make the following assumption on π which is required, in addition to Assumption B, to ensure that \mathcal{L} possesses a spectral gap in $L_0^2(\pi)$.

Assumption C

$$\lim_{|x| \rightarrow +\infty} |\nabla \log \pi(x)| = \infty. \tag{19}$$

In the following lemma we establish a number of fundamental properties relating to the spectrum of \mathcal{L} . Using these results, we then establish the well-posedness of the Poisson Eq. (13) for any $f \in L_0^2(\pi)$.

Lemma 2 *Suppose that Assumptions B and C hold. Then the embedding $\mathcal{H}^1 \subset L_0^2(\pi)$ is compact. For all $\alpha \in \mathbb{R}$, the operator \mathcal{L} satisfies the following Poincaré inequality in $L_0^2(\pi)$:*

$$\lambda \|g\|_\pi^2 \leq \langle g, (-\mathcal{L})g \rangle_\pi, \quad g \in \mathcal{H}^1, \tag{20}$$

where λ is a positive constant, independent of the nonreversible perturbation. Moreover, for all $f \in L_0^2(\pi)$ there exists a unique $\phi \in \mathcal{H}^1$ such that $-\mathcal{L}\phi = f$.

Proof Assumption B clearly implies that

$$-\Delta \log \pi(x) \leq (1 - \delta) |\nabla \log \pi(x)|^2 + M_1, \quad x \in \mathbb{R}^d, \tag{21}$$

for some $\delta \in (0, 1)$ and $M_1 > 0$. Applying [40, Theorem 8.5.3], relations (19) and (21) imply the compactness of the embedding $\mathcal{H}^1 \subset L_0^2(\pi)$. As a result, following [40, Theorem 8.6.1], this is sufficient for the Poincaré inequality to hold for \mathcal{S} , i.e.

$$\lambda \|g\|_\pi^2 \leq \|g\|_1^2 = \langle g, (-\mathcal{S})g \rangle_\pi, \quad g \in \mathcal{H}^1.$$

from which (20) follows by the antisymmetry of \mathcal{A} in $L_0^2(\pi)$. The existence of this Poincaré inequality implies that the semigroup P_t^γ associated with (6) converges exponentially fast to equilibrium, that is, for all $f \in L_0^2(\pi)$:

$$\|P_t^\gamma f\|_\pi \leq e^{-\lambda t} \|f\|_\pi, \quad t \geq 0. \tag{22}$$

Given $f \in L_0^2(\pi)$, define

$$\phi(x) = \int_0^\infty P_s^\gamma f(x) ds. \tag{23}$$

By (22) it follows that $\phi \in L_0^2(\pi)$ and from (23) we have $-\mathcal{L}\phi = f$. Moreover, we have the bound

$$\|\nabla \phi\|_\pi^2 = \langle \phi, (-\mathcal{L})\phi \rangle_\pi = \langle f, \phi \rangle_\pi,$$

so that

$$\|\phi\|_1 \leq C \|f\|_\pi < \infty.$$

□

Throughout this section, we shall assume that Assumptions B and C hold, and moreover, for simplicity we shall make the following additional assumption:

Assumption D The nonreversible perturbation γ is smooth and bounded in L^∞ .

We believe that this assumption could be relaxed, see in particular [30] for more general assumptions under which the following results should hold. We stick here to a simple presentation. In practice, this assumption is not very stringent. Indeed, suppose that there exists a smooth function $\psi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ such that

$$\psi(V(x)) |\nabla V(x)| \leq 1, \quad \text{for all } x \in \mathbb{R}^d, \tag{24}$$

where $V = -\log \pi$. Define

$$\gamma = J \nabla V \psi(V), \tag{25}$$

then γ satisfies (7) since

$$\begin{aligned} \nabla \cdot (\pi \gamma) &= -\frac{1}{Z} \nabla \cdot \left(J \nabla e^{-V} \psi(V) \right) \\ &= -\frac{1}{Z} \nabla \cdot \left(J \nabla e^{-V} \right) \psi(V) - \frac{1}{Z} \psi'(V) \nabla V \cdot J \nabla V e^{-V} \\ &= 0, \end{aligned}$$

using the fact that J is antisymmetric. Moreover, it is clear that $\gamma(x)$ is bounded and smooth, thus satisfying Assumption D. In particular, if V has compact level sets (which holds for example if V is a nonnegative polynomial function), then condition (24) is satisfied by choosing ψ to be a smooth non-negative function with compact support.

We note that γ of the form (25) leaves the potential V invariant under the flow $\dot{z}_t = \gamma(z_t)$, i.e. $V(z_t)$ is constant for all $t \geq 0$. Thus, for large $|\alpha|$, the flow $\alpha\gamma$ will result in rapid exploration of the level surfaces of V , but the motion of X_t between level surfaces is entirely due to the reversible dynamics of the process. In particular, for potentials with energy barriers, the transition time for X_t^γ to cross a barrier will still satisfy the same Arrhenius law as the corresponding reversible process. Other choices of flow γ are possible. For example, one could alternatively consider a skew-symmetric matrix function $J(x)$ as detailed in [41]. The corresponding flow would then be defined by

$$\gamma(x) = -J(x)\nabla V(x) + \nabla \cdot J(x), \quad x \in \mathbb{R}^d \tag{26}$$

It is straightforward to check that $\nabla \cdot (\gamma\pi) = 0$, using the fact that $J(x)$ is skew-symmetric. If additionally, the matrix function J is smooth with bounded derivative and compact support, then γ satisfies Assumption D. As detailed in [41] one can further generalise this choice of dynamics by additionally introducing a space dependent diffusion tensor, and an appropriate correction of the drift to maintain ergodicity with respect to π . We do not consider this choice of dynamics in this paper, noting that most of the presented results can be readily generalized to this scenario.

Given that Assumption D holds for all function $u \in \mathcal{H}^1$, then we have

$$\begin{aligned} \|\mathcal{A}u\|_\pi &\leq \|\gamma\|_{L^\infty(\mathbb{R}^d)} \|u\|_1 \text{ and } \int \gamma(x) \cdot \nabla u(x)\pi(dx) \\ &= - \int \nabla \cdot (\gamma(x) \cdot \pi(x)) u(x) dx = 0, \end{aligned}$$

by (7), so that the operator $\mathcal{A} : \mathcal{H}^1 \rightarrow L_0^2(\pi)$ defined by $\mathcal{A} = \gamma \cdot \nabla$ is well defined.

3.2 An Expression for the Asymptotic Variance

The main objective of this paper is to study the effect of the nonreversible perturbation γ on the asymptotic variance σ_f^2 , with the aim of choosing γ so that σ_f^2 is minimized. Integrating (16) by parts we can express σ_f^2 in terms of the Dirichlet form for \mathcal{L} as follows:

$$\begin{aligned} \frac{1}{2}\sigma_f^2 &= \int |\nabla\phi(x)|^2 \pi(dx) \\ &= - \int \phi(x)\nabla \cdot (\nabla\phi(x)\pi(x)) dx \\ &= - \int \phi(x) (\Delta\phi(x) + \nabla \log \pi(x) \cdot \nabla\phi(x)) \pi(x) dx \\ &= \langle \phi, (-\mathcal{S})\phi \rangle_\pi \\ &= \langle \phi, (-\mathcal{L})\phi \rangle_\pi, \end{aligned} \tag{27}$$

where the last line follows from the fact that \mathcal{A} is antisymmetric in $L_0^2(\pi)$.

Starting from (27) we can obtain a quite explicit characterisation of σ_f^2 . Indeed, given $f \in L_0^2(\pi)$, we can rewrite the last line of (27) as

$$\sigma_f^2 = 2\langle f, (-\mathcal{L})^{-1} f \rangle_\pi = 2\left\langle f, [(-\mathcal{L})^{-1}]^S f \right\rangle_\pi, \tag{28}$$

where $[\cdot]^S$ denotes the symmetric part of the operator. Using the fact that $-\mathcal{L}$ is invertible on $L_0^2(\pi)$ for all $\alpha \in \mathbb{R}$ by Lemma 2, the following result yields an expression for $\sigma_f^2(\alpha)$ in terms of \mathcal{S} , \mathcal{A} and α .

Lemma 3 *Let $\mathcal{L} = S + \alpha\mathcal{A}$, for $\alpha \in \mathbb{R}$. Then we have*

$$[(-\mathcal{L})^{-1}]^S = [-S + \alpha^2\mathcal{A}^*(-S)^{-1}\mathcal{A}]^{-1}, \tag{29}$$

where $\mathcal{A}^* = -\mathcal{A}$ denotes the adjoint of \mathcal{A} in $L_0^2(\pi)$. In particular, for all $f \in L_0^2(\pi)$,

$$\sigma_f^2(\alpha) = 2\langle f, (-\mathcal{L})^{-1}f \rangle_\pi \leq 2\langle f, (-S)^{-1}f \rangle_\pi = \sigma_f^2(0). \tag{30}$$

Proof Since $-S$ is positive, the operator $\mathcal{Q} = (-S)^{-\frac{1}{2}} : L_0^2(\pi) \rightarrow \mathcal{H}^1$ can be defined using functional calculus. Consider the operator $\mathcal{C} := \mathcal{Q}\mathcal{A}$. We can write $-S + \alpha^2\mathcal{A}^*(-S)^{-1}\mathcal{A} = -S + \alpha^2\mathcal{C}^*\mathcal{C}$, which can be shown to be closed in $L_0^2(\pi)$ with domain \mathcal{H}^2 . Moreover, this operator has nullspace $\{0\}$, so that the inverse is also densely defined on $L_0^2(\pi)$. To show that (29) holds, we expand the left hand side to get

$$\begin{aligned} [(-\mathcal{L})^{-1}]^S &= \frac{1}{2} [(-S + \alpha\mathcal{A})^{-1} + (-S - \alpha\mathcal{A})^{-1}] \\ &= \frac{1}{2} \mathcal{Q} [(I + \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} + (I - \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1}] \mathcal{Q}. \end{aligned}$$

Since

$$\begin{aligned} (I + \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} + (I - \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} &= (I + \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} (I - \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} (I - \alpha\mathcal{Q}\mathcal{A}\mathcal{Q}) \\ &\quad + (I - \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} (I + \alpha\mathcal{Q}\mathcal{A}\mathcal{Q})^{-1} (I + \alpha\mathcal{Q}\mathcal{A}\mathcal{Q}) \\ &= [I - \alpha^2\mathcal{Q}\mathcal{A}\mathcal{Q}^2\mathcal{A}\mathcal{Q}]^{-1} 2I, \end{aligned}$$

therefore

$$[(-\mathcal{L})^{-1}]^S = \mathcal{Q} [I - \alpha^2\mathcal{Q}\mathcal{A}\mathcal{Q}^2\mathcal{A}\mathcal{Q}]^{-1} \mathcal{Q} = [-S + \alpha^2\mathcal{A}^*(-S)^{-1}\mathcal{A}]^{-1},$$

as required. The inequality (30) is then simply a consequence of the fact that

$$-S + \alpha^2\mathcal{A}^*(-S)^{-1}\mathcal{A} \succeq -S$$

where \succeq denotes the partial ordering between selfadjoint operators on $L_0^2(\pi)$. □

Thus, the asymptotic variance is never increased by introducing a nonreversible perturbation, for all $f \in L_0^2(\pi)$. This had already been noted in [61] where an expression for the the asymptotic was derived as the curvature of the rate function of the empirical measure, and also in [30] using an approach similar to that above. Expression (28) provides us with a formula for σ_f^2 in terms of a symmetric quadratic form which is explicit in terms of \mathcal{A} and S .

3.3 Quantitative Estimates for the Asymptotic Variance

In this section we derive quantitative versions of (29) and (30), using techniques developed in [58] for the analysis of the Green–Kubo formula, which is itself based on earlier work on the estimation of the eddy diffusivity in turbulent diffusion [2, 8, 44, 57].

Following the approach of [22, 58], to quantify the effect of the antisymmetric perturbation $\alpha\mathcal{A}$ on the asymptotic variance, we define the operator $\mathcal{G} = (-S)^{-1}\mathcal{A} : \mathcal{H}^1 \rightarrow \mathcal{H}^1$. First we note that we can rewrite (29) as follows:

$$\begin{aligned} [(-\mathcal{L})^{-1}]^S &= [-S + \alpha^2\mathcal{A}^*(-S)^{-1}\mathcal{A}]^{-1} \\ &= [I + \alpha^2(-S)^{-1}\mathcal{A}^*(-S)^{-1}\mathcal{A}]^{-1} (-S)^{-1} \\ &= [I - \alpha^2\mathcal{G}^2]^{-1} (-S)^{-1} \end{aligned} \tag{31}$$

and therefore, from (28) and (29):

$$\begin{aligned} \frac{1}{2}\sigma_f^2(\alpha) &= \left\langle (-S)\hat{f}, [I - \alpha^2\mathcal{G}^2]^{-1}(-S)^{-1}(-S)\hat{f} \right\rangle_\pi \\ &= \left\langle \hat{f}, (-S)[I - \alpha^2\mathcal{G}^2]^{-1}\hat{f} \right\rangle_\pi \\ &= \left\langle \hat{f}, [I - \alpha^2\mathcal{G}^2]^{-1}\hat{f} \right\rangle_1. \end{aligned} \tag{32}$$

From the boundedness of the nonreversible perturbation we have the following properties:

Lemma 4 *Suppose that Assumption D holds, then the operator $\mathcal{G} = (-S)^{-1}A$ is skew-adjoint on \mathcal{H}^1 .*

Proof For $f, g \in \mathcal{H}^1$:

$$\langle \mathcal{G}f, g \rangle_1 = \langle (-S)^{-1}Af, (-S)g \rangle_\pi = -\langle f, Ag \rangle_\pi = -\langle f, (-S)(-S)^{-1}Ag \rangle_\pi = -\langle f, \mathcal{G}g \rangle_1,$$

so that \mathcal{G} is antisymmetric. Since \mathcal{G} is also bounded, it follows that \mathcal{G} is skewadjoint on \mathcal{H}^1 . \square

Moreover, under appropriate assumptions on the target distribution π , one can show that the operator \mathcal{G} is compact.

Lemma 5 *Suppose that Assumptions B, C and D hold, then the operator \mathcal{G} is compact on \mathcal{H}^1 .*

Proof By Lemma 2, the embedding $\mathcal{H}^1 \subset L_0^2(\pi)$ is compact, and it follows immediately that $\mathcal{H}^2 \subset \mathcal{H}^1$ is a compact embedding. Moreover, since γ is bounded we have,

$$\|\mathcal{G}f\|_2^2 = \langle (-S)\mathcal{G}f, (-S)\mathcal{G}f \rangle_\pi = \langle Af, Af \rangle_\pi \leq \|f\|_1^2, \quad f \in \mathcal{H}^1,$$

from which the result follows. \square

Extending \mathcal{H}^1 to its complexification, there exists a compact selfadjoint operator Γ on \mathcal{H}^1 such that $\mathcal{G} = i\Gamma$. From the spectral theorem for compact selfadjoint operators [24, Theorem 6.21], the eigenfunctions of Γ form a complete orthonormal basis in \mathcal{H}^1 , and the eigenvalues of Γ are real. We can partition the eigenfunctions into those spanning $\mathcal{N} := \text{Ker}[\Gamma] = \text{Ker}[A]$ and those spanning \mathcal{N}^\perp . We denote by

$$\{\lambda_n\}_{n=1,2,3,\dots}, \quad \text{and} \quad \{e_n\}_{n=1,2,3,\dots},$$

the eigenvalues and corresponding eigenfunctions of the operator Γ restricted to \mathcal{N}^\perp . For $f \in L_0^2(\pi)$ we have the following unique decomposition for $\hat{f} = (-S)^{-1}f$ in \mathcal{H}^1 :

$$\hat{f} = \hat{f}_\mathcal{N} + \sum_{n=1}^\infty \hat{f}_n e_n, \tag{33}$$

where $\hat{f}_\mathcal{N} \in \mathcal{N}$ and $\hat{f}_n = \langle \hat{f}, e_n \rangle_1$. Consequently, using this spectral decomposition, we can rewrite (32) as

$$\begin{aligned} \sigma_f^2(\alpha) &= 2 \left\langle \hat{f}, [I - \alpha^2\mathcal{G}^2]^{-1}\hat{f} \right\rangle_1 \\ &= 2 \left\langle \hat{f}, [I - \alpha^2(i\Gamma)^2]^{-1}\hat{f} \right\rangle_1 \\ &= 2 \left\langle \hat{f}, [I + \alpha^2\Gamma^2]^{-1}\hat{f} \right\rangle_1 \\ &= 2 \left\| \hat{f}_\mathcal{N} \right\|_1^2 + 2 \sum_{n=1}^{+\infty} \frac{1}{1 + \alpha^2\lambda_n^2} \left| \hat{f}_n \right|^2. \end{aligned} \tag{34}$$

The conclusion of the above computation is summarized in the following result.

Theorem 4 *Suppose that Assumptions B, C and D hold. Let $f \in L_0^2(\pi)$ and $\alpha \in \mathbb{R}$, then the asymptotic variance $\sigma_f^2(\alpha)$ corresponding to X_t^γ is given by (34). In particular, we obtain the following limiting values for the asymptotic variance:*

$$\lim_{\alpha \rightarrow 0} \sigma_f^2(\alpha) = \sigma_f^2(0) = 2\|\hat{f}\|_1^2, \tag{35}$$

and,

$$\lim_{\alpha \rightarrow \pm\infty} \sigma_f^2(\alpha) = 2\|\hat{f}_{\mathcal{N}}\|_1^2. \tag{36}$$

□

From (34) we obtain the following bounds on the asymptotic variance

$$\sigma_f^2(\alpha) = 2\|\hat{f}_{\mathcal{N}}\|_1^2 + 2\sum_{n=1}^{+\infty} \frac{|\hat{f}_n|^2}{1 + \alpha^2\lambda_n^2} \leq 2\|\hat{f}_{\mathcal{N}}\|_1^2 + 2\sum_{n=1}^{+\infty} |\hat{f}_n|^2 = 2\|\hat{f}\|_1^2 = \sigma_f^2(0).$$

The problem of choosing the optimal nonreversible perturbation in (6) to minimize the asymptotic variance over all observables in $L_0^2(\pi)$ can be expressed as the following min-max problem:

$$\min_{\gamma \in \mathbb{A}_M} \max_{f \in L_0^2(\pi)} \frac{\langle f, (-\mathcal{S} - \mathcal{A}(-\mathcal{S})^{-1}\mathcal{A})^{-1}f \rangle_\pi}{\|f\|_\pi^2}, \quad \text{where } \mathcal{A} = \gamma \cdot \nabla,$$

where, for some constant $M > 0$, \mathbb{A}_M denotes the set of admissible nonreversible drifts, typically,

$$\mathbb{A}_M = \left\{ \gamma \in C_b^\infty(\mathbb{R}^d; \mathbb{R}^d) : \nabla \cdot (\gamma\pi) = 0 \quad \text{and} \quad \|\gamma\|_{L^\infty} < M \right\}.$$

This is equivalent to finding $\gamma \in \mathbb{A}_M$ which solves this max-min problem

$$\max_{\gamma \in \mathbb{A}_M} \min \sigma(-\mathcal{S} + \mathcal{A}^*(-\mathcal{S})^{-1}\mathcal{A}), \tag{37}$$

where $\sigma[\mathcal{M}]$ denotes the spectrum of the operator \mathcal{M} on $L_0^2(\pi)$. It is important to make the distinction between this problem and that considered in [37] for finding the optimal spectral gap, namely finding $\gamma \in \mathbb{A}_M$ such that

$$\max_{\gamma \in \mathbb{A}_M} \min \text{Re}(\sigma(-\mathcal{S} - \mathcal{A})). \tag{38}$$

From (37) and (38) it is evident that the decrease in asymptotic variance, and the increase in spectral gap arising from a nonreversible perturbation are due to very different mechanisms. For the latter problem we see that the increase of spectral gap arises from the nonnormality of $\mathcal{S} + \mathcal{A}$. In addition, since the operator $\mathcal{A}^*(-\mathcal{S})^{-1}\mathcal{A}$ is nonnegative in \mathcal{H}^1 , a nonreversible perturbation cannot increase the variance. This is in contrast with the problem of maximising the speed of convergence to equilibrium, as was considered in [37], where increasing the strength of the nonreversible perturbation could result in a decrease of the speed of convergence.

We note however that a nonreversible perturbation γ which solves the min-max problem (37) need not be a good candidate for reducing the variance of $\pi_T(f)$ for a given fixed observable $f \in L_0^2(\pi)$. Indeed, unless $\mathcal{N} = \{0\}$ for all $g \in \mathcal{H}^1$, there will always be an observable for which the nonreversible perturbation does not reduce the variance, since if $g \in \mathcal{N}$ is nonzero, $f = (-\mathcal{S})g$ is nonzero and such that $\sigma_f^2(\alpha) = \sigma_f^2(0)$ for any α . Thus,

from a practical point of view, it makes more sense to consider the problem choosing γ to minimise the asymptotic variance of $\pi_T(f)$ for a particular observable. To this end, for a fixed $f \in L^2_0(\pi)$, we can identify two distinct cases: $(-S)^{-1}f \in \mathcal{N}^\perp$, in which case

$$\lim_{|\alpha| \rightarrow \infty} \sigma_f^2(\alpha) = 0,$$

and $(-S)^{-1}f \notin \mathcal{N}^\perp$ in which case,

$$\lim_{|\alpha| \rightarrow \infty} \sigma_f^2(\alpha) = 2 \left\| \hat{f}_{\mathcal{N}} \right\|_1^2 > 0.$$

More generally, consider $f \in L^2(\pi)$ so that $f - \pi(f) \in L^2_0(\pi)$. Assuming we can increase α arbitrarily, and neglecting any computational issues arising from the resulting discretisation error (which will be discussed in Sect. 6), the problem of minimising the asymptotic variance of f reduces to finding γ such that

$$(-S)^{-1}(f - \pi(f)) \perp \mathcal{N}, \tag{39}$$

holds in \mathcal{H}^1 . Checking this condition requires the solution of an elliptic boundary value problem, thus this condition is not of practical use. Nonetheless, we can derive some intuition from (39). Clearly, if we can choose γ so that $\mathcal{N} = \{0\}$ in \mathcal{H}^1 , the nonreversible perturbation will be optimal for all observables $f \in L^2(\pi)$. In general, it might not be possible to find γ that satisfies this condition. For the nonreversible drift considered in [37], namely $\gamma = J\nabla V$, with $J^\top = -J$, \mathcal{N} will always be nontrivial; indeed, in this case $H \circ V \in \mathcal{N}$ for all functions H such that $H \circ V \in \mathcal{H}^1$. In this case, it is always possible to choose an observable f such that γ will not be optimal for f , in the sense that $\sigma_f^2(\alpha)$ is nonzero in the limit $\alpha \rightarrow \infty$. We also remark that these asymptotic results, in particular the distinction between these two cases is reminiscent of similar results that have been obtained in the context of turbulent diffusion [44, 57].

An analogous classification of the asymptotic behaviour of the spectral gap of the operator (9) in the limit of large α is considered in [18] on a compact manifold. Indeed, in [18, Theorem 1] it is determined that the spectral gap is finite in the limit of $\alpha \rightarrow \pm\infty$ if and only if \mathcal{A} has a nonconstant eigenfunction in \mathcal{H}^1 . However, one should note that, the asymptotic variance $\sigma_f^2(\alpha)$ may converge to 0 as $\alpha \rightarrow \pm\infty$ even when the spectral gap is finite, see also [61, Example 2.9] for a counterexample.

3.4 A Two Dimensional Example

In this section we present a simple example on \mathbb{R}^2 . Consider the problem of calculating $\pi(f)$ with respect to the Gaussian distribution $\pi(x) = \frac{1}{2}e^{-|x|^2/2}$. The reversible overdamped Langevin equation corresponding to π is given by the OU process:

$$dX_t = -X_t dt + \sqrt{2} dW_t. \tag{40}$$

We introduce an antisymmetric perturbation $\alpha\gamma(X_t)$ where $\gamma(x)$ is the flow given by

$$\gamma(x) = \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix}.$$

The infinitesimal generator for the perturbed nonreversible dynamics is

$$\mathcal{L} = \mathcal{S} + \alpha\mathcal{A},$$

where

$$Sf(x) = -x \cdot \nabla f(x) + \Delta f(x) \quad \text{and} \quad \mathcal{A}f(x) = \gamma(x) \cdot \nabla f(x).$$

In polar coordinates, this is given by

$$\begin{aligned} \mathcal{L}f(r, \theta) &= \left(-r + \frac{1}{r}\right) \partial_r f(r, \theta) + \partial_{r,r} f(r, \theta) + \alpha \partial_\theta f(r, \theta) \\ &\quad + \frac{1}{r^2} f_{\theta,\theta}(r, \theta), \quad (r, \theta) \in \mathbb{R}_{>0} \times (0, 2\pi). \end{aligned}$$

As $|\alpha| \rightarrow +\infty$, we expect the deterministic flow to move increasingly along the level curves of the distribution $\pi(x)$. Given an observable f , any variance in $\pi_T(f)$ arising from the variation of f along the level curves should vanish as $|\alpha| \rightarrow \infty$, leaving only the variance contributed by the variation of f between level curves. We make this precise with a particular example. Consider the observable $f(x_1, x_2) = 2x_1^2$, expressible in polar coordinates as

$$f(r, \theta) = 2r^2 \cos^2(\theta) = r^2 (1 + \cos(2\theta)).$$

Noting that $\pi(f) = 2$, it is straightforward to check that the Poisson equation $-\mathcal{L}\phi = f - \pi(f)$, has mean zero solution given by

$$\phi(r, \theta) = \left(\frac{r^2}{2} - 1\right) + \frac{r^2 (\cos(2\theta) - \alpha \sin(2\theta))}{1 + \alpha^2}$$

The asymptotic variance can be evaluated directly as follows

$$\begin{aligned} \sigma_f^2 &= 2 \langle \phi, (-\mathcal{L})\phi \rangle_\pi \\ &= \frac{2}{Z} \int_0^{2\pi} \int_0^\infty \left(\left(\frac{r^2}{2} - 1\right) + \frac{r^2 (\cos(2\theta) - \alpha \sin(2\theta))}{1 + \alpha^2} \right) (r^2(1 + \cos(2\theta)) - 2) r e^{-r^2/2} dr d\theta \\ &= 4 \left(1 + \frac{1}{1 + \alpha^2} \right). \end{aligned}$$

Therefore, as $|\alpha| \rightarrow \infty$, the asymptotic variance converges to 4. We note that, in this case,

$$\hat{f} = (-S)^{-1} (f - \pi(f)) = \left(\frac{r^2}{2} - 1\right) + \frac{r^2}{2} (\cos(2\theta)),$$

where $(r^2/2 - 1)$ is perpendicular to $r^2/2 \cos(2\theta)$ in \mathcal{H}^1 . Since the nullspace of \mathcal{A} consists of all \mathcal{H}^1 functions which depend only on r , we have $\hat{f}_{\mathcal{N}} = r^2/2 - 1$. It follows that $2 \|\hat{f}_{\mathcal{N}}\|_1^2 = 4$, so that

$$\lim_{|\alpha| \rightarrow \infty} \sigma_f^2(\alpha) = 2 \|\hat{f}_{\mathcal{N}}\|_1^2,$$

which agrees with the conclusions of Theorem 4. As $|\alpha| \rightarrow \infty$, the motion in the θ direction is averaged out. Indeed, in this limit, $2 \|\hat{f}_{\mathcal{N}}\|_1^2$ corresponds to the asymptotic variance of the observable $\int_0^T r_t^2 dt$, where r_t is the following 1D reversible process,

$$dr_t = \left(-r_t + \frac{1}{r_t}\right) dt + \sqrt{2} dW_t, \quad r_0 > 0.$$

For more general observables $f(r, \theta)$, we expect maximum variance reduction as $|\alpha| \rightarrow \infty$ when f varies strongly with respect to θ . In the other extreme, we expect zero improvement

when f depends only on r . This intuition is formalised in the following section, in particular in Proposition 1 and the subsequent bound (51).

For more general potentials, using a flow field of the form $\gamma(x) = J \nabla \log \pi(x)$, $J^\top = -J$, the mechanism for reducing the asymptotic variance is analogous: the large antisymmetric drift gives rise to fast deterministic mixing along the level curves of the potential, while the reversible dynamics induce slow diffusive motion along the gradient of the potential. When the nullspace of \mathcal{A} is trivial in \mathcal{H}^1 , the fast deterministic flow is ergodic, so that, for α large, the antisymmetric component will cause a rapid exploration of the entire state space. Consequently, the asymptotic variance converges to 0 as $\alpha \rightarrow \infty$. On the other hand, if \mathcal{A} has a nontrivial nullspace, the antisymmetric perturbation is no longer ergodic, and the state space can be decomposed into components such that the rapid flow behaves ergodically in each individual component. In the limit of large α , X_t becomes a fast-slow system, with rapid exploration within the ergodic components coupled to a slow diffusion between components. Very recently, Rey-Bellet and Spiliopoulos [62] have applied Freidlin–Wenzell theory to rigorously analyse this case in the large α limit for a large class of potentials.

4 Nonreversible Perturbations of Gaussian Diffusions

For the case when the target distribution is Gaussian, the SDE (6) for X_t^γ is linear. In this case, we can obtain an explicit analytical expression for the asymptotic variance for a large class of observables f . Indeed, consider the nonsymmetric Ornstein-Uhlenbeck process in \mathbb{R}^d :

$$dX_t^\gamma = -(I + \alpha J)X_t^\gamma dt + \sqrt{2} dW_t, \tag{41}$$

where J is an antisymmetric matrix, $\alpha > 0$, and W_t is a standard d -dimensional Brownian motion. The stationary distribution $\pi(x)$ is $\mathcal{N}(0, I)$, independent of α and J . Although this system does not fall under the framework of Theorem 4 we are still able to obtain analogous conditions for a reduction in the asymptotic variance. The objective of this section is to mirror the results for speeding up convergence to equilibrium of X_t that were derived in [37] to the case of minimizing the asymptotic variance. In particular, following arguments similar to [58, Sect. 4.2], an explicit formula for the asymptotic variance will be derived, from which an optimal J can be chosen, in a manner similar to [37]. We note that for the process (41), the optimal nonreversible perturbation obtained in [37] does not provide any increase to the rate of convergence to equilibrium, since all eigenvalues of the covariance matrix of the Gaussian stationary distribution are the same. Nonetheless, in this section we show that for certain observables, the asymptotic variance of $\pi_T(f)$ can be dramatically decreased.

4.1 Explicit Formula for the Asymptotic Variance

We shall assume that the observable f is a quadratic functional of the form

$$f(x) = x \cdot Mx + l \cdot x + k, \tag{42}$$

where $M \in \mathbb{R}^{d \times d}$ is a symmetric positive definite matrix, $l \in \mathbb{R}^d$ and k is a constant, chosen so that $f(x)$ is centered with respect to $\pi(x)$: $k = -\text{Tr } M$. Consider the Poisson Eq. (10):

$$-\mathcal{L}\phi(x) = x \cdot Mx + l \cdot x - \text{Tr } M, \quad \pi(\phi) = 0, \tag{43}$$

where \mathcal{L} is the infinitesimal generator of (41) given by $\mathcal{L} = -(I + \alpha J)x \cdot \nabla + \Delta$. For such observables we can solve the Poisson Eq. (43) analytically and obtain a closed-form formula for the asymptotic variance for the observable f .

Proposition 1 *Let $A = (I + \alpha J)^\top = (I - \alpha J)$. The unique mean zero solution of the Poisson Eq. (43) is given by*

$$\phi(x) = x \cdot Cx + D \cdot x - \text{Tr}(C),$$

where

$$C = \frac{1}{2} \int_0^\infty e^{-As} M e^{-A^\top s} ds, \tag{44}$$

and

$$D = A^{-1}l.$$

Moreover, the asymptotic variance is given by

$$\sigma_f^2(\alpha) = \|M\|_F^2 - \int_0^\infty e^{-2s} \left\| [M, e^{-\alpha J s}] \right\|_F^2 ds + 2l \cdot (I + \alpha^2 J^\top J)^{-1}l,$$

where $\|A\|_F = \sqrt{\text{Tr}[AA^\top]}$ denotes the Frobenius norm of A , and $[A, B] = AB - BA$ is the commutator of A and B . In particular,

$$\sigma_f^2(\alpha) \leq \sigma_f^2(0), \quad \forall \alpha \in \mathbb{R}. \tag{45}$$

Proof Clearly $\phi(x)$ must also be a quadratic function of x , so we make the ansatz

$$\phi(x) = x \cdot Cx + D \cdot x - \text{Tr} C.$$

Plugging this into (43) we obtain

$$x \cdot A \left((C + C^\top)x + D \right) - \text{Tr}(C + C^\top) = x \cdot Mx + l \cdot x - \text{Tr} M.$$

Comparing equal powers of x we have

$$x \cdot A(C + C^\top)x = x \cdot Mx, \tag{46a}$$

$$AD \cdot x = l \cdot x \tag{46b}$$

$$\text{Tr} C = \frac{1}{2} \text{Tr} M \tag{46c}$$

for all $x \in \mathbb{R}^d$. Since x is arbitrary, it follows that

$$D = A^{-1}l.$$

Equation (46a) is a Lyapunov equation, which is well-posed as M is positive definite and $\text{spec}(A) \subset \{\lambda \in \mathbb{C} : \text{Re} \lambda > 0\}$. Indeed, for C given by

$$C = \frac{1}{2} \int_0^\infty e^{-As} M e^{-A^\top s} ds,$$

both (46a) and (46c) are satisfied. The asymptotic variance $\sigma_f^2(\alpha)$ is then given by (see (27)),

$$\begin{aligned} \frac{1}{2}\sigma_f^2(\alpha) &= \langle \phi(x), f(x) \rangle_\pi = \langle x \cdot Cx + D \cdot x - \text{Tr } C, x \cdot Mx + l \cdot x - \text{Tr } M \rangle_\pi \\ &= \langle x \cdot Cx, x \cdot Mx \rangle_\pi + \langle D \cdot x, l \cdot x \rangle_\pi \\ &\quad - \langle \text{Tr } C, x \cdot Mx \rangle_\pi - \langle x \cdot Cx, \text{Tr } M \rangle_\pi + \text{Tr } C \text{Tr } M. \end{aligned} \tag{47}$$

Using the fact that C and M are symmetric,

$$\langle x \cdot Cx, x \cdot Mx \rangle_\pi = \sum_{i,j,k,l} C_{ij}M_{kl}\mathbb{E}_\pi [x_i x_j x_k x_l] = \text{Tr } C \text{Tr } M + 2 \text{Tr}(CM^\top).$$

Therefore,

$$\begin{aligned} \frac{1}{2}\sigma_f^2(\alpha) &= 2 \text{Tr}(CM^\top) + \langle D \cdot x, l \cdot x \rangle_\pi \\ &= 2 \text{Tr}(CM^\top) + D \cdot l \\ &= \int_0^\infty \text{Tr} \left[e^{-As} M e^{-A^\top s} M^\top \right] ds + l \cdot A^{-1}l. \end{aligned} \tag{48}$$

Since $A = I - \alpha J$, where J is antisymmetric, we have that, for all $l \in \mathbb{R}^d$:

$$\begin{aligned} l \cdot A^{-1}l &= \frac{1}{2}l \cdot [(I - \alpha J)^{-1} + (I + \alpha J)^{-1}]l \\ &= \frac{1}{2}l \cdot [(I - \alpha J)^{-1}(I + \alpha J)^{-1}(I + \alpha J) + (I + \alpha J)^{-1}(I - \alpha J)^{-1}(I - \alpha J)]l \\ &= l \cdot (I + \alpha^2 J^\top J)^{-1}l. \end{aligned}$$

Therefore, we obtain the following expression for $\sigma_f^2(\alpha)$:

$$\sigma_f^2(\alpha) = 2 \int_0^\infty e^{-2s} \text{Tr} \left[e^{\alpha Js} M e^{-\alpha Js} M^\top \right] ds + 2l \cdot (I + \alpha^2 J^\top J)^{-1}l. \tag{49}$$

Writing the first term as follows:

$$\int_0^\infty e^{-2s} \text{Tr} \left[e^{\alpha Js} M e^{-\alpha Js} M^\top \right] ds = \frac{1}{2}\|M\|_F^2 - \frac{1}{2} \int_0^\infty e^{-2s} \left\| \left[M, e^{-\alpha Js} \right] \right\|_F^2 ds.$$

From this, and the facts that $J^\top J \geq 0$ and $[M, I] = 0$, the inequality (45) follows. □

Since J is skew-symmetric, the matrix exponential $e^{\alpha Js}$ is a rotation matrix. Thus, the matrix $e^{\alpha Js} M e^{-\alpha Js}$ has the same eigenvalues as M and M^\top . From [4, III. 6.14] we have

$$\lambda^\downarrow(M) \cdot \lambda^\uparrow(M) = \lambda^\downarrow(e^{\alpha Js} M e^{-\alpha Js}) \cdot \lambda^\uparrow(M^\top) \leq \text{Tr} \left[e^{\alpha Js} M e^{-\alpha Js} M^\top \right],$$

where $\lambda^\uparrow(M)$ and $\lambda^\downarrow(M)$ denote the vectors of eigenvalues of M , sorted in ascending and descending order, respectively. In particular,

$$\int_0^\infty e^{-2s} \text{Tr} \left[e^{\alpha Js} M e^{-\alpha Js} M^\top \right] ds \geq \frac{1}{2}\lambda^\downarrow(M) \cdot \lambda^\uparrow(M). \tag{50}$$

On the other hand, let $\mathcal{N} = \mathcal{N}[J] = \mathcal{N}[J^\top J]$, and diagonalising $J^\top J$ we obtain

$$J^\top J = U^\top D U, \quad \text{where } D = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \alpha \Lambda \end{bmatrix}, \quad \text{for } \Lambda = \text{diag}(\lambda_1, \dots, \lambda_{d-N}).$$

where U is a $d \times d$ orthogonal matrix with the first N columns spanning \mathcal{N} . Then

$$l \cdot (I + \alpha^2 J^\top J)^{-1} l = U l \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & (\mathbf{I} + \alpha \Lambda)^{-1} \end{bmatrix} U l \xrightarrow{\alpha \rightarrow \pm\infty} U l \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} U l = |l_{\mathcal{N}}|^2,$$

where $l_{\mathcal{N}}$ is the projection of l onto \mathcal{N} . Thus, for quadratic observables, from (49), the asymptotic variance has the following lower bound in the limit of large α :

$$\lim_{\alpha \rightarrow \infty} \sigma_f^2(\alpha) \geq \underline{\sigma}_f^2 := \lambda^\downarrow(M) \cdot \lambda^\uparrow(M) + 2 |l_{\mathcal{N}}|^2, \tag{51}$$

In particular, for diffusions with linear drift, the asymptotic variance cannot be decreased arbitrarily. The worst case scenario is when $f(x) = m(|x|^2 - d)$, for some $m > 0$, for which the asymptotic variance will not be decreased, for any antisymmetric matrix J and $\alpha \in \mathbb{R}$, analogous to the situation which occurs in [37] for maximising the spectral gap, when all the eigenvalues of the covariance matrix are equal.

4.2 Finding the Optimal Perturbation

We first focus on the case when the observable is linear, i.e. $f(x) = l \cdot x$, similar to that considered in [58, Sect. 4.2], so that $\sigma_f^2 = 2l \cdot (I + \alpha^2 J^\top J)^{-1} l$. Suppose that we fix $\alpha \in \mathbb{R}$, and wish to choose J which minimizes the asymptotic variance subject to $\|J\|_F = 1$. In this case, we have the equality

$$\lim_{\alpha \rightarrow \pm\infty} \sigma_f^2 = 2 |l_{\mathcal{N}}|^2,$$

which is optimal if l is orthogonal to \mathcal{N} . Thus, the best we can do is to choose J such that l is an eigenvector of $J^\top J$ with maximal eigenvalue. This can be done by choosing a unit vector $\omega \in \mathbb{R}^d$ orthogonal to l and setting

$$J = \frac{(\tilde{l} \otimes \omega - \omega \otimes \tilde{l})}{\sqrt{2}},$$

where $\tilde{l} = l/|l|$. Then,

$$J^* J = \frac{1}{2} (-\tilde{l} \otimes \omega + \omega \otimes \tilde{l}) (\tilde{l} \otimes \omega - \omega \otimes \tilde{l}) = \frac{\tilde{l} \otimes \tilde{l} + \omega \otimes \omega}{2},$$

is a projector onto $\{l, \omega\}$ and $\|J\|_F = 1$, where $\|\cdot\|_F$ is the Frobenius norm with respect to the Euclidean basis. In this case

$$\sigma_f^2(\alpha) = 2l \cdot (I + \alpha^2 J^\top J)^{-1} l = \frac{4|l|^2}{2 + \alpha^2}.$$

We note in addition that, when minimising the asymptotic variance, there are many antisymmetric matrices J which give the minimal asymptotic variance. As an example, let $d = 3$, consider the observables $f(x) = l^{(i)} \cdot x$ for $l^{(1)} = (0, 1, 1)^\top/\sqrt{2}$ and $l^{(2)} = (1, 0, 1)^\top/\sqrt{2}$ and $l^{(3)} = (1, -1, 1)^\top/\sqrt{3}$, respectively. We choose J to be

$$J = \frac{1}{\sqrt{6}} \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}. \tag{52}$$

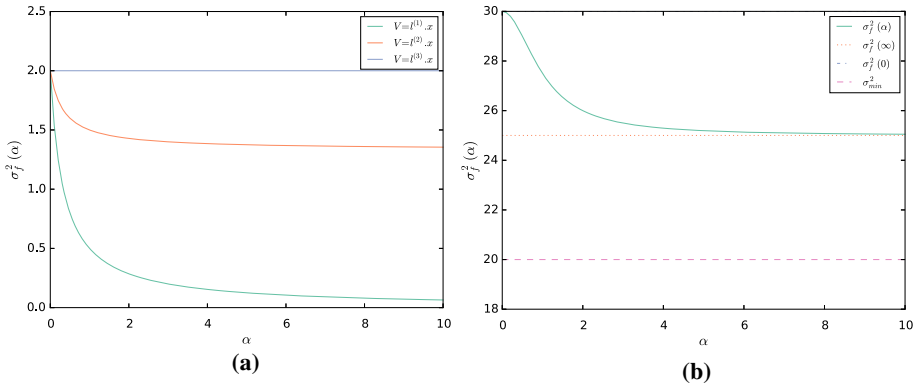


Fig. 2 The asymptotic variance for a linear diffusion X_t given in (41) for linear observables $f(x) = l^{(i)} \cdot x$, $i = 1, 2, 3$ (left), and for a quadratic observable $f(x) = x \cdot M_1 x$ (right) (Color figure online)

In this case

$$\sigma_f^2(\alpha) = \frac{2}{6 + 3\alpha^2} l \cdot \begin{pmatrix} 6 + \alpha^2 & -\alpha^2 & \alpha^2 \\ -\alpha^2 & 6 + \alpha^2 & -\alpha^2 \\ \alpha^2 & -\alpha^2 & 6 + \alpha^2 \end{pmatrix} l.$$

The decay of $\sigma_f^2(\alpha)$ as $\alpha \rightarrow \infty$ is strongly dependent on the nullspace of J , given by

$$\mathcal{N} = \text{span}[\zeta], \quad \text{where } \zeta = (1, -1, 1).$$

In Fig. 2a we plot the asymptotic variance for these observables. For $f(x) = l^{(1)} \cdot x$, see that as $\alpha \rightarrow \infty$ the asymptotic variance converges to 0. This is due to the fact that $l^{(1)}$ is perpendicular to the nullspace of J . Thus, for this observable the matrix J given by (52) is an optimal perturbation. For $f(x) = l^{(2)} \cdot x$, since $l^{(2)}$ is not orthogonal to ζ , as $\alpha \rightarrow \infty$,

$$\sigma_f^2(\alpha) \rightarrow 2 \left| l_{\mathcal{N}}^{(2)} \right|^2 = \frac{4}{3}.$$

Finally, for $f(x) = l^{(3)} \cdot x$ we observe that the asymptotic variance remains constant at 2, since $l^{(3)} \in \mathcal{N}$, the nonreversible perturbation has no effect on this observable.

We now focus on the case when $l = 0$ so that $f(x) = x \cdot Mx - \text{Tr}(M)$. In this case, it is not clear how to construct an optimal J , however the bound (51) suggests a good candidate for J . Suppose that M has eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ with corresponding eigenvectors e_1, \dots, e_d . Suppose that d is even. Let

$$\{i_1, j_1\}, \dots, \{i_{d/2}, j_{d/2}\},$$

be a partition of $1, \dots, d$ into disjoint pairs. Define the antisymmetric matrix J by

$$J = \sum_{k=1}^{d/2} e_{i_k} \otimes e_{j_k} - e_{j_k} \otimes e_{i_k}, \quad k = 1, \dots, \frac{d}{2}. \tag{53}$$

In this case $e^{\alpha J s}$ can be decomposed into a product of rotations between the pairs of eigenvectors:

$$e^{\alpha J s} = \prod_{k=1}^{d/2} R_{e_{i_k}, e_{j_k}}(\alpha s),$$

where $R_{v,w}(\theta)$ is an anticlockwise rotation of angle θ in the $\{v, w\}$ plane. Then

$$\int_0^\infty e^{-2s} \text{Tr} \left[e^{\alpha J s} M e^{-\alpha J s} M^\top \right] ds = \frac{1}{4(1 + \alpha^2)} \sum_{k=1}^{d/2} (\lambda_{i_k} - \lambda_{j_k})^2 + \frac{1}{4} \sum_{k=1}^{d/2} (\lambda_{i_k} + \lambda_{j_k})^2.$$

For this choice of J , the asymptotic variance is given by

$$\lim_{\alpha \rightarrow \pm\infty} \sigma_f^2(\alpha) = \frac{1}{2} \sum_{k=1}^{d/2} (\lambda_{i_k} + \lambda_{j_k})^2.$$

Arguing by contradiction, this is minimized by choosing $i_k = k$, and $j_k = (d - k + 1)$ for $k = 1, \dots, d/2$, in which case

$$\begin{aligned} \lim_{\alpha \rightarrow \pm\infty} \sigma_f^2(\alpha) &= \frac{1}{2} [(\lambda_1 + \lambda_d)^2 + (\lambda_2 + \lambda_{d-1})^2 + \dots + (\lambda_{d/2} + \lambda_{d/2+1})^2] \\ &= \frac{1}{2} \lambda^\downarrow(M) \cdot \lambda^\uparrow(M) + \frac{1}{2} \sum_{k=1}^d \lambda_k^2. \end{aligned}$$

Clearly,

$$\lambda^\downarrow(M) \cdot \lambda^\uparrow(M) \leq \frac{1}{2} \lambda^\downarrow(M) \cdot \lambda^\uparrow(M) + \frac{1}{2} \sum_{k=1}^d \lambda_k^2 \leq \text{Tr}(M^2),$$

however these bounds are only tight when $M = m I$, for $m > 0$.

As an example, on Fig. 2b we plot the asymptotic variance for a quadratic observable $x \cdot M_1 x$ for a linear diffusion (41), where $d = 4$. The matrix M_1 is given by

$$M_1 = \begin{pmatrix} 3/2 & -1/2 & 0 & 0 \\ -1/2 & 3/2 & 0 & 0 \\ 0 & 0 & 7/2 & -1/2 \\ 0 & 0 & -1/2 & 7/2 \end{pmatrix},$$

with eigenvalues $\lambda_i = i$, for $i = 1, \dots, 4$. When $\alpha = 0$, the asymptotic variance is $\sigma_f^2(0) = 30$. The lower bound (45) is given by $\underline{\sigma}_f^2 = 2(\lambda_1 \lambda_4 + \lambda_2 \lambda_3) = 20$. We choose the antisymmetric matrix $J : \mathbb{R}^{4 \times 4}$ as in (53), that is

$$J = \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The resulting asymptotic variance $\sigma_f^2(\alpha)$ is plotted as a function of α in Fig. 2a which converges to $\frac{1}{2}((\lambda_1 + \lambda_4)^2 + (\lambda_2 + \lambda_3)^2) = 25$ as $\alpha \rightarrow \infty$.

5 Numerical Experiments

In this section we present three numerical examples illustrating the effects of the antisymmetric perturbation on the asymptotic variance σ_f^2 . We will consider target measures defined

by Gibbs distributions of the form:

$$\pi(x) = \frac{1}{Z} e^{-\beta V(x)}, \tag{54}$$

where $V(x)$ is a given smooth, confining potential with finite unknown normalisation constant Z . We will also assume that the divergence-free vector field $\gamma(x)$ is given by

$$\gamma(x) = -\alpha J \nabla V(x), \quad J = -J^\top, \tag{55}$$

so that the drift of (6) will be given by

$$b(x) = -(\beta I + \alpha J) \nabla V(x).$$

The symmetric and antisymmetric parts of the generator in $L^2(\mathbb{T}^d; \pi(x)dx)$ are, respectively,

$$\mathcal{S} = -\beta \nabla V \cdot \nabla + \Delta, \quad \mathcal{A} = J \nabla V \cdot \nabla.$$

Provided that $V \in H^1(\pi)$ the nullspace of the antisymmetric part of the generator is always nonempty. Indeed, the antisymmetry of \mathcal{A} implies that $\mathcal{A}V = 0$. Thus, for any observable f such that $\pi(f) = 0$ and

$$0 \neq \langle f, V \rangle_\pi,$$

since $\langle \hat{f}, V \rangle_1 = \langle (-\mathcal{S})^{-1} f, (-\mathcal{S})V \rangle_\pi = \langle f, V \rangle_\pi$ the conclusion of Theorem 4 implies the asymptotic variance $\sigma_f^2(\alpha)$ will converge to a nonzero constant as $\alpha \rightarrow \infty$.

5.1 Periodic Distribution

In the first example we consider a two-dimensional target density given by (54) where the periodic potential $V(x)$ is given by

$$V(x) = \sin(2\pi x_1) \cos(2\pi x_2), \quad x = (x_1, x_2) \in \mathbb{T}^2, \tag{56}$$

and $\beta^{-1} = 0.1$. Our objective is to calculate the expectation $\pi(f)$ of the observable

$$f(x) = 1 + 4 \sin(4\pi x_1)^2 + 4 \cos(4\pi x_2)^2. \tag{57}$$

In Fig. 3a we plot the asymptotic variance over $\alpha \in [0, 12]$. The stepsize Δt for the Euler–Maruyama discretisation of (6) is chosen to be 10^{-3} . For each α , $M = 10^3$ independent realisations of the Markov process are run, starting from $X_0 = (0, 0)$, each for $T = 10^5$ time units to ensure that the process is close to stationarity. We observe that increasing α from 0 to 10 decreases the asymptotic variance by approximately two orders of magnitude. In Fig. 3b we plot the value of the estimator along with the confidence intervals estimated over 10^3 independent realisations. For this particular example, it appears that increasing the magnitude of the nonreversible perturbation does not appear to give rise to any noticeable increase in bias, while it significantly decreases the asymptotic variance, giving rise to a dramatic increase in performance of the estimator $\pi_T(f)$.

5.2 Warped Gaussian Distribution

As a second numerical example we consider computing an observable with respect to a two-dimensional warped Gaussian distribution [23], defined by (54) with

$$V(x) = \frac{x_1^2}{100} + (x_2 + bx_1^2 - 100b)^2. \tag{58}$$

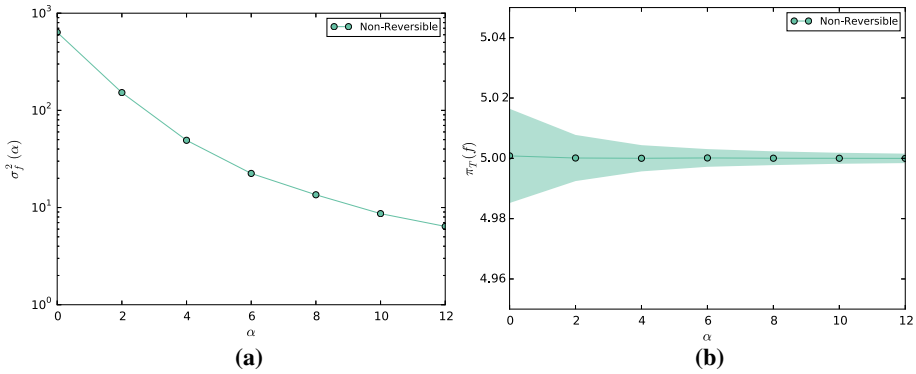


Fig. 3 The asymptotic variance and value of the estimator $\pi_T(f)$, where $T = 10^5$, for the target Gibbs distribution with potential defined by (56) and observable as in (57). The plot was generated from 10^3 independent realisations of an Euler–Maruyama discretisation of the nonreversible diffusion (6), with stepsize 10^{-3} over $T = 10^5$ time-units. The shaded region in **b** indicates the 95 % confidence interval of the estimator for the given α (Color figure online)

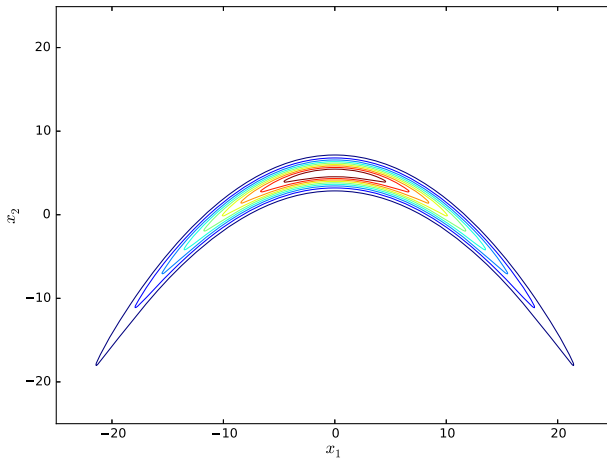


Fig. 4 Contour plot of the warped Gaussian distribution with potential (58)

The parameter $b > 0$ is chosen to be $b = 0.05$. The potential $V(x)$ is plotted in Fig. 4. Our objective is to compute $\pi(f)$ where the observable f is given by:

$$f(x) = x_1^2 + x_2^2.$$

The nonreversible perturbation is chosen to be:

$$\gamma(x) = -J\nabla V(x), \quad \text{where } J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{59}$$

In Fig. 5a we plot the asymptotic variance of the estimator $\pi_T(f)$, approximated from 10^3 independent realisations of the process, each run for 10^8 timesteps of size 10^{-3} so that the total time is $T = 10^5$. Each realisation was started at $(0, 0)$. As expected, we observe a significant decrease in asymptotic variance as the magnitude of the nonreversible perturbation is increased. Increasing α from 0 to 10 gives rise to a decrease in variance by a factor of

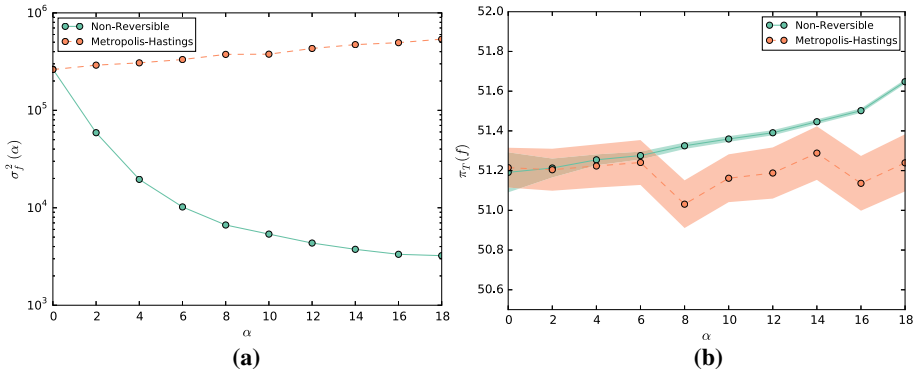


Fig. 5 Plot of the asymptotic variance $\sigma_f^2(\alpha)$ and the estimator $\pi_T(f)$ for the warped Gaussian distribution π defined by (58) and the observable $f(x) = |x|^2$ for different values of α . The solid line was generated from 10^3 independent realisations of an Euler–Maruyama discretisation of the nonreversible diffusion (6), with stepsize 10^{-3} over $T = 10^5$ time-units. The shaded region in b indicates the confidence interval of the estimator for the given α . The dashed line denotes the variance and mean when the chain is augmented with a accept/reject step, see Sect. 5.3 (Color figure online)

80. In Fig. 5b we plot the value of the estimator π_T averaged over the 10^3 realisations, with corresponding confidence intervals. As α is increased, the bias arising from the discretisation error also increases. To mitigate this bias one could decrease the stepsize, thus requiring more steps to generate $\pi_T(f)$ or otherwise introduce a Metropolis–Hastings accept–reject step, which will be discussed in Sect. 5.3. The tradeoff between bias and variance is considered in more detail in Sect. 6.

5.3 Introducing a Metropolis–Hastings Accept–Reject Step

When sampling from a distribution π it is natural to introduce a Metropolis–Hastings (MH) accept–reject step, rather than use the Markov chain obtained directly from an Euler–Maruyama discretization of the SDE (6). Given a current state $X^{(n)}$, a next state is proposed according to the Euler–Maruyama discretisation of (6):

$$\tilde{X} \sim \mathcal{N}\left(X^{(n)} + \Delta t \nabla \log \pi(X^{(n)}), \Delta t\right), \tag{60}$$

The proposed state is then accepted (i.e. $X^{(n+1)} := \tilde{X}$) with probability

$$r(X^{(n)}, \tilde{X}) = 1 \wedge \frac{\pi(\tilde{X})p(X^{(n)}, \tilde{X})}{\pi(X^{(n)})p(\tilde{X}, X^{(n)})},$$

where $p(\cdot, x)$ is the proposal density corresponding to (60). By introducing this accept–reject mechanism, the resulting chain $X^{(n)}$ is guaranteed to have a stationary distribution which is exactly π , independently of Δt . Thus, introducing a Metropolis–Hastings accept–reject step allows for far larger stepsizes to be used, while still preserving the correct invariant distribution of the chain, eliminating any bias arising from the discretisation of the SDE, making it beneficial both in terms of computational performance and stability.

A natural question is whether an MH chain using a proposal distribution based on the SDE (6) with antisymmetric drift will inherit the superior mixing properties of the nonreversible

diffusion process. As the MH algorithm works by enforcing the detailed balance of the chain $X^{(n)}$ with respect to the distribution π , we expect that any benefits of the antisymmetric drift term will be negated when introducing this accept–reject step. To test this, we repeat the numerical experiment of Sect. 5.2 using the MH algorithm using the Euler discretisation of (6) as a proposal scheme, for various values of α . The effect of introducing this accept–reject step to the nonreversible diffusion is evident from Fig. 5. While the accept–reject step removes any bias due to discretisation error, as is evident from Fig. 5b, the asymptotic variance actually increases as α increases. This is due to the fact that for large α , proposals are more likely to be rejected as they are far away from the current state.

5.4 Dimer in a Solvent

We now test the effect of adding a nonreversible perturbation to a test model from molecular dynamics, as described in [38, Sect. 1.3.2.4]. We consider a system composed of N particles P_1, \dots, P_N in a two-dimensional periodic box of side length L . Particles P_1 and P_2 are assumed to form a dimer pair, in a solvent comprising the particles P_3, \dots, P_N . The solvent particles interact through a truncated Lennard-Jones potential:

$$V_{WCA}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] + \epsilon, & \text{if } r \leq r_0, \\ 0 & \text{if } r > r_0, \end{cases}$$

where r is the distance between two particles, ϵ and σ are two positive parameters, and $r_0 = 2^{\frac{1}{6}}\sigma$. The interaction potential between the dimer pair is given by a double-well potential

$$V_S(r) = h \left[1 - \frac{(r - r_0 - w)^2}{w^2} \right]^2, \tag{61}$$

where h and w are two positive parameters. The total energy of the system is given by

$$V(q) = V_S(|q_1 - q_2|) + \sum_{3 \leq i < j \leq N} V_{WCA}(|q_i - q_j|) + \sum_{i=1,2} \sum_{3 \leq j \leq N} V_{WCA}(|q_i - q_j|), \tag{62}$$

with $q = (q_1, \dots, q_N) \in (L\mathbb{T})^{dN}$, where q_i denotes the position of particle P_i . The potential V_S has two energy minima at $r = r_0$ (corresponding to the compact state), and at $r = r_0 + 2w$ (corresponding to the stretched state). The energy barrier separating the two states is h . Define $\xi(q)$ to be

$$\xi(q) = \frac{|q_1 - q_2| - r_0}{2w}.$$

This reaction coordinate describes the transition from the compact state, $\xi(q) = 0$ to the stretched state $\xi(q) = 1$. The standard overdamped reversible dynamics to sample from the stationary distribution $\pi(q) \propto \exp(-\beta V(q))$ is given by

$$dq_t = -\nabla V(q_t) dt + \sqrt{2\beta^{-1}} dW_t,$$

where β is the inverse temperature. Our objective is to compute the average reaction coordinate $\mathbb{E}_\pi \xi(q)$. We introduce an antisymmetric drift term as follows

$$dq_t = -(I + \alpha J)\nabla V(q_t) dt + \sqrt{2\beta^{-1}} dW_t, \tag{63}$$

where $J \in \mathbb{R}^{2N \times 2N}$ is an antisymmetric matrix. We consider two types of nonreversible perturbations, determined by antisymmetric matrices, J_1 and J_2 . The first matrix J_1 is the block-circulant matrix defined by:

$$J_1 = \begin{pmatrix} O_2 & I_2 & \cdots & O_2 & -I_2 \\ -I_2 & O_2 & I_2 & & O_2 \\ \vdots & -I_2 & O_2 & \ddots & \vdots \\ O_2 & & \ddots & \ddots & I_2 \\ I_2 & O_2 & \cdots & -I_2 & O_2 \end{pmatrix},$$

where I_2 and O_2 denote the 2×2 identity and zero matrix, respectively. The second matrix is given by

$$J_2 = \begin{pmatrix} R & O_4 & \cdots & O_4 \\ O_4 & O_4 & \cdots & O_4 \\ \vdots & \vdots & \ddots & \vdots \\ O_4 & O_4 & \cdots & O_4 \end{pmatrix},$$

where R is the following rotation matrix on $\mathbb{R}^{4 \times 4}$:

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

and O_4 is the 4×4 zero matrix. The effect, in this case, is to apply the antisymmetric transformation only on the first two coordinates, which correspond to the positions of the particles composing the dimer.

Using an Euler–Maruyama discretisation of (63), samples are generated for $N = 16$ particles, with parameter values $\beta = 1.0$, $\sigma = 1.0$, $\epsilon = 1.0$, $h = 1.0$, $w = 0$, and $\alpha \in \{0, 2, 4, 6, 8, 10\}$. For each value of α , a single realisation, starting from $\mathbf{0}$ is simulated for 10^{10} timesteps of size $\Delta t = 10^{-5}$, so that the total running time is $T = 10^5$. The asymptotic variance is approximated from a single realisation of the process using a batch-means estimator (see [1, IV .5]). In Fig. 6 we plot the value of the generated estimator, along with the asymptotic variance for different values of α and for J_1 and J_2 . The first perturbation provides marginally lower asymptotic variance for this particular observable, giving rise to a 66 % decrease, as opposed to a 50 % decrease for J_2 , increasing α from 0 to 10. This decrease in asymptotic variance, comes at the cost of increased computation time due to having to reduce the stepsize to compensate for the discretisation error and numerical instability caused by the large drift. While this may appear as a negative result, since the antisymmetric drift terms were chosen arbitrarily, no claim is made about optimality. An important issue is whether the increase in computational cost outweighs the benefits. This issue will be studied more carefully in Sect. 6.

6 The Computational Cost of Nonreversible Langevin Samplers

As observed Sect. 5, while increasing α is guaranteed to decrease the asymptotic variance $\sigma_f^2(\alpha)$ for the estimator $\pi_T(f)$, this will also give rise to an increase in the discretisation error arising from the particular discretisation being used. Moreover, as α increases, the SDE

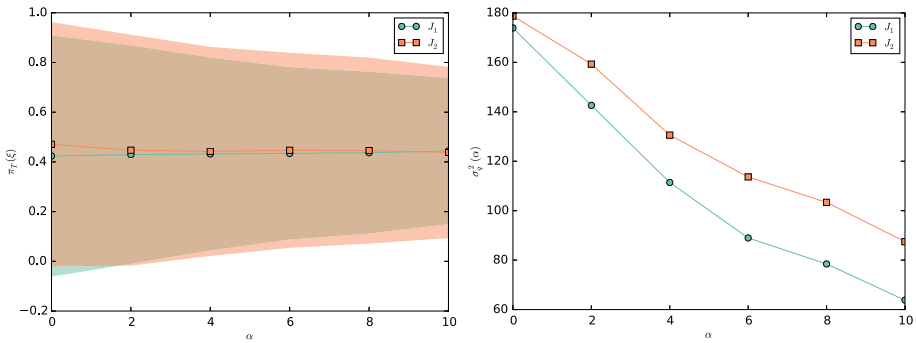


Fig. 6 Value of the estimator $\pi_T(\xi)$ for $T = 10^5$ with corresponding 95 % confidence intervals, for the dimer model, generated from a single realisation of (63). The *right* plot shows the asymptotic variance of the estimator over varying α , with antisymmetric drifts defined by $J_1 \nabla V(x)$ and $J_2 \nabla V(x)$, respectively (Color figure online)

(6) becomes more and more stiff, to the extent that the discretisation becomes numerically unstable unless the stepsize is chosen to be accordingly small. As a result, any discretisation $\{X^{(n)}\}_{n=1}^N$ will require smaller timesteps to guarantee that the stationary distribution of $X^{(n)}$ is sufficiently close to $\pi(x)$. This tradeoff between computational cost and asymptotic variance of the estimator must be taken into consideration when comparing reversible to nonreversible diffusions.

A rigorous error analysis of the long time average estimator was carried out in [47]. In this paper careful estimates of the mean square error

$$Err_{N, \Delta t}^2(f) := \mathbb{E} \left| \frac{1}{N} \sum_{n=1}^N f(X^{(n)}) - \pi(f) \right|^2$$

were derived for discretisations of a general overdamped Langevin diffusion on the unit torus \mathbb{T}^d , see also [70]. In particular, in [47, Theorem 5.2] it is shown that the mean squared error can be bounded as follows

$$Err_{N, \Delta t}^2[f] \leq C \left(\Delta t^2 + \frac{1}{N \Delta t} \right), \tag{64}$$

where C is a positive constant independent of Δt and N , which depends on the coefficients of the SDE and the observable f . This estimate makes explicit the tradeoff between discretisation error and sampling error. For a fixed computational budget N , the right hand side of (64) is minimized when $\Delta t \propto (N)^{-\frac{1}{3}}$. For an SDE of the form (6), we expect that the constant C will increase with α . Identifying the correct scaling of the error with respect to α is an interesting problem that we intend to study.

To obtain a clearer idea of the bias variance tradeoff we compute the mean-square error for the Euler–Maruyama discretisation for two particular examples. In Fig. 7a, we consider the warped Gaussian distribution defined by (58) and the observable $f(x) = |x|^2$. A value for $\pi(f)$ is obtained by integrating $\int_{\mathbb{R}^d} f(x)\pi(dx)$ numerically, using a globally adaptive quadrature scheme to obtain an approximation with error less than 10^{-12} . In Fig. 7a we plot the relative mean-squared-error defined by $(Err_{N, \Delta t}[f]/\pi(f))^2$ for an Euler–Maruyama discretisation of (6), for timestep Δt in the interval $[2^{-5}, 1]$. The total number of timesteps is kept fixed at $N = 10^6$. For each value of α , the mean square error is approximated over an

ensemble of 256 independent realisations. Missing points indicate finite time blowup of the discretized diffusion. The dashed line denotes the MSE generated from the corresponding MALA sampler, namely an Euler–Maruyama discretisation of the reversible diffusion with an added Metropolis–Hastings accept–reject step. We note that both the Euler–Maruyama discretisation and the MALA sampler require one evaluation of the gradient term $\nabla \log \pi$ per timestep, so that comparing an ergodic average obtained from 10^6 steps of each scheme is fair.

A trade-off between discretisation error and variance is evident from Fig. 7a, and is consistent with the error estimate (64). We observe that the nonreversible Langevin sampler outperforms the reversible Langevin sampler by an order of magnitude, with the lowest MSE attained when $\alpha = 10$. As α is increased beyond this point, the discretisation error is balanced by the decrease in variance, and we observe no further gain in performance. Nonetheless, despite the fact that the MALA scheme has no bias, the nonreversible sampler, with $\alpha = 5$, outperforms MALA (in terms of MSE) by a significant factor of 8.8.

We repeat this numerical experiment for the target distribution π given by a standard Gaussian distribution in \mathbb{R}^d and observable $f(x) = x_2 + x_3$. In this case $\pi(f)$ is exacty 0. We use the linear diffusion specified by (41) where the antisymmetric matrix J is given in (52), which is optimal for this observable. We plot the (absolute) MSE for the estimator $\pi_T(f)$ in Fig. 8a. While the smallest MSE is attained by the nonreversible Langevin sampler, when $\alpha = 25$, the increase in performance is only marginal. This is due to the fact that increasingly smaller timesteps must be taken to ensure that the EM approximation does not blow up. Indeed, the $\alpha = 25$ sampler would not converge to a finite value for Δt greater than 10^{-3} , while the reversible sampler ($\alpha = 0$) and the MALA scheme were accurate even for timesteps of order 1.

From both examples it is clear that managing the numerical stability and discretisation error of the skew-symmetric drift term is essential for any practical implementation of the nonreversible sampling scheme. This suggests that using higher order and/or more stable numerical integrators to compute long time averages would be beneficial to eliminate the bias arising from discretisation, as well as permit the use of larger timesteps. Naturally, such schemes would require multiple evaluations of $\nabla \log \pi$ at each step, thus it is possible that the additional computational cost offsets any performance gain. In the remainder of this section, we will perform the same numerical experiments using an integrator based on a Strang splitting [35,67] of the stochastic reversible and deterministic nonreversible dynamics. The reversible part will be simulated using a standard MALA scheme and the nonreversible flow using an appropriate higher–order integrator. Indeed, denote by $\Phi_{r,t}$ the evolution of the reversible SDE (3) from time 0 to t , and let $\Phi_{n,t}(x)$ denote the flow map corresponding to the ODE:

$$\dot{z}(t) = \gamma(z(t)), \quad z(0) = x.$$

We shall consider an integrator based on the following map from time t to $t + \Delta t$:

$$\Psi_{\Delta t} = \Phi_{r,\Delta t/2} \circ \Phi_{n,\Delta t} \circ \Phi_{r,\Delta t/2}.$$

For this implementation, we approximate $\Phi_{r,\Delta t}(x)$ using a single step of a MALA scheme with proposal based on (3) with stepsize Δt . The nonreversible flow $\Phi_{n,\Delta t}$ is approximated using a fourth-order Runge–Kutta method.

We leave the justification and analysis of this scheme as the goal of future work, and in this paper simply use it to compute a long time average approximation to $\pi(f)$ and compare the MSE with that of a corresponding reversible MALA scheme. To obtain a fair comparison

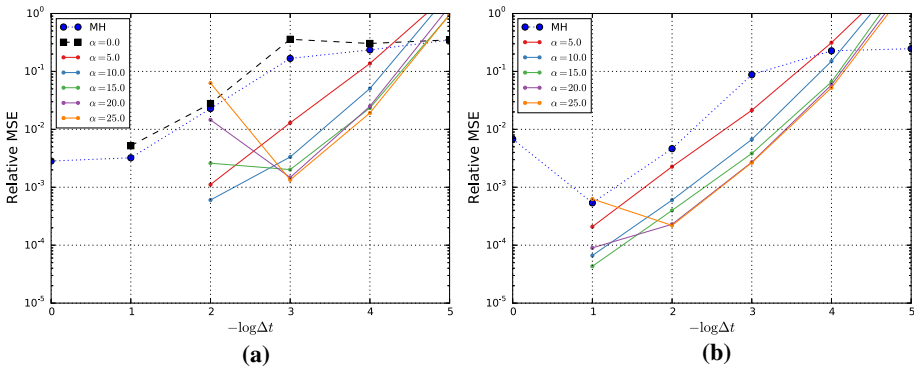


Fig. 7 Plot of the relative mean-square error as a function of the timestep size Δt for the observable $f(x) = |x|^2$ of a warped Gaussian distribution defined by (58), with a fixed computational budget of 10^6 evaluations of $\nabla \log \pi$ for (a) and $6 \cdot 10^6$ for (b). The dotted line denotes the relative MSE for the MALA scheme (Color figure online)

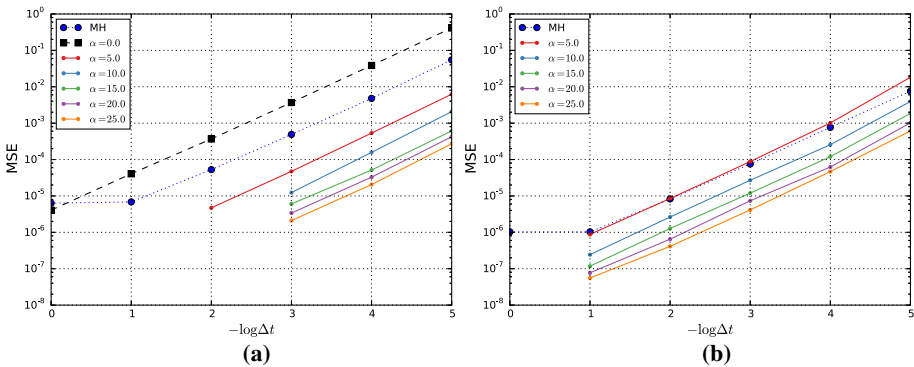


Fig. 8 Plot of the mean-square error as a function of the timestep size Δt for a linear observable of a Gaussian distribution, so that X_t^γ is a linear diffusion process. A fixed computational budget of 10^6 evaluations of $\nabla \log \pi$ is imposed in (a), and $6 \cdot 10^6$ in (b). The dotted line depicts the corresponding error for the metropolized sampler (Color figure online)

between the results obtained by MALA and the splitting scheme, we note that while a careful implementation of MALA requires only one evaluation of $\nabla \log \pi$ per timestep, the splitting scheme requires six evaluations of $\nabla \log \pi$ per timestep (naively a single timestep would require two evaluations for each reversible substep and four evaluations for the nonreversible substep, however we can reuse two evaluations of $\nabla \log \pi$ between the steps). Thus, we shall compare the MSE obtained from trajectories of 10^6 timesteps of the nonreversible sampler with $6 \cdot 10^6$ timesteps of the corresponding MALA scheme, for stepsizes ranging from 10^{-5} to 1. The results for the warped Gaussian distribution in \mathbb{R}^2 and standard Gaussian in \mathbb{R}^3 are plotted in Figs. 7b and 8b, respectively. Note that we omit the $\alpha = 0$ case since, in this case, the splitting scheme reduces to standard the MALA scheme. We observe that with this splitting scheme, the nonreversible sampler outperforms MALA by a factor of 13 for the warped Gaussian model, and by a factor of 20 for the standard Gaussian model. The benefits of the splitting scheme appear to be twofold: firstly the integrator is more stable, in both models, the long time simulation of X_t^γ did not blow up, even for large values of α , and for

$\Delta t = 0.1$. Moreover, compared to the corresponding Euler–Maruyama discretisation, the MSE is consistently an order of magnitude less.

While this splitting scheme is only a first step into properly investigating appropriate integrators for nonreversible Langevin schemes, the above numerical experiments demonstrate clearly that there is a significant benefit in doing so, which motivates future investigation.

7 Conclusions and Further Work

In this paper we have presented a detailed analytical and numerical study of the effect of nonreversible perturbations to Langevin samplers. In particular, we have focused on the effect on the asymptotic variance of adding a nonreversible drift to the overdamped Langevin dynamics. Our theoretical analysis, presented for diffusions with periodic coefficients and for diffusions with linear drift for which a complete analytical study can be performed, and our numerical investigations on toy models clearly show that a judicious choice of the nonreversible drift can lead to a substantial reduction in the asymptotic variance. On the other hand, as observed from the dimer model example in Sect. 5.4, an arbitrary choice of nonreversible drift will not always give rise to significant improvement. We have also presented a careful study of the computational cost of the algorithm based on a nonreversible Langevin sampler, in which the competing effects of reducing the asymptotic variance and of increasing the stiffness due to the addition of a nonreversible drift are monitored. The main conclusions that can be drawn from our numerical experiments is that a nonreversible Langevin sampler with close-to-the-optimal choice of the nonreversible drift significantly outperforms the (reversible) Metropolis–Hastings sampler.

There are many open problems that need to be studied further:

- (1) The effect of using degenerate, hypoelliptic diffusions for sampling from a given distribution.
- (2) Combining the use of nonreversible Langevin samplers with standard variance reduction techniques such as the zero variance reduction MCMC methodology [14].
- (3) Optimizing Langevin samplers within the class of reversible diffusions.
- (4) The development of nonreversible Metropolis–Hastings algorithms based on the above techniques, possibly related to approach described in [9].
- (5) The development and analysis of numerical schemes specifically designed to simulate nonreversible Langevin diffusions.

All these topics are currently under investigation.

Acknowledgments The work of TL is supported by the European Research Council under the European Union’s Seventh Framework Program/ERC Grant Agreement Number 614492. GP thanks Ch. Doering for useful discussions and comments. The research of AD is supported by the EPSRC under Grant No. EP/J009636/1. GP is partially supported by the EPSRC under Grants Nos. EP/J009636/1, EP/L024926/1, EP/L020564/1 and EP/L025159/1. TL would like to thank J. Roussel for pointing out some mistakes in a preliminary version of the manuscript. The authors would like to thank the anonymous referees for their useful comments and suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Asmussen, S., Glynn, P.W.: *Stochastic Simulation: Algorithms and Analysis*, vol. 57. Springer, New York (2007)
2. Avellanada, M., Majda, A.J.: Stieltjes integral representation and effective diffusivity bounds for turbulent transport. *Phys. Rev. Lett.* **62**, 753–755 (1989)
3. Avellanada, M., Majda, A.J.: An integral representation and bounds on the effective diffusivity in passive advection by laminar and turbulent flows. *Commun. Math. Phys.* **138**(2), 339–391 (1991)
4. Bhatia, R.: *Matrix Analysis*, vol. 169. Springer, New York (1997)
5. Bhattacharya, R.: A central limit theorem for diffusions with periodic coefficients. *Ann. Probab.* **13**(2), 385–396 (1985)
6. Bhattacharya, R.N.: On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **60**(2), 185–201 (1982)
7. Bhattacharya, R.N.: Multiscale diffusion processes with periodic coefficients and an application to solute transport in porous media. *Ann. Appl. Probab.* **9**(4), 951–1020 (1999)
8. Bhattacharya, R.N., Gupta, V.K., Walker, H.F.: Asymptotics of solute dispersion in periodic porous media. *SIAM J. Appl. Math.* **49**(1), 86–98 (1989)
9. Bierkens, J.: Nonreversible Metropolis-Hastings. [arXiv:1401.8087](https://arxiv.org/abs/1401.8087) (2014)
10. Bovier, A., Eckhoff, M., Gaynard, V., Klein, M.: Metastability in reversible diffusion processes: sharp asymptotics for capacities and exit times. Technical report, WIAS, Berlin (2002)
11. Bovier, A., Gaynard, V., Klein, M.: Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc.* **7**(1), 69–99 (2005)
12. Cattiaux, P., Chafaï, D., Guillin, A.: Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA* **9**(2), 337–382 (2012)
13. Constantin, P., Kiselev, A., Ryzhik, L., Zlato, A.: Diffusion and mixing in fluid flow. *Ann. Math.* **168**(2), 643–674 (2008)
14. Dellaportas, P., Kontoyiannis, I.: Control variates for estimation based on reversible Markov chain Monte Carlo samplers. *J. R. Stat. Soc. B* **74**(1), 133–161 (2012)
15. Diaconis, P., Holmes, S., Neal, R.M.: Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10**(3), 726–752 (2000)
16. Douc, R., Fort, G., Guillin, A.: Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stoch. Processes Appl.* **119**(3), 897–923 (2009)
17. Down, D., Meyn, S.P., Tweedie, R.L.: Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23**(4), 1671–1691 (1995)
18. Franke, B., Hwang, C.R., Pai, H.M., Sheu, S.J.: The behavior of the spectral gap under growing drift. *Trans. Am. Math. Soc.* **362**(3), 1325–1350 (2010)
19. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: *Bayesian Data Analysis*, vol. 2. Taylor & Francis, Boca Raton (2014)
20. Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. B* **73**(2), 123–214 (2011)
21. Glynn, P.W., Meyn, S.P.: A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24**(2), 916–931 (1996)
22. Golden, K., Papanicolaou, G.: Bounds for effective parameters of heterogeneous media by analytic continuation. *Commun. Math. Phys.* **90**(4), 473–491 (1983)
23. Haario, H., Saksman, E., Tamminen, J.: Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Stat.* **14**(3), 375–396 (1999)
24. Helffer, B.: *Spectral Theory and Its Applications*, vol. 139. Cambridge University Press, Cambridge (2013)
25. Helland, I.S.: Central limit theorems for martingales with discrete or continuous time. *Scand. J. Stat.* **9**(2), 79–94 (1982)
26. Hennequin, G., Aitchison, L., Lengyel, M.: Fast sampling-based inference in balanced neuronal networks. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems 27*, pp. 2240–2248. Curran Associates, Inc., New York (2014)
27. Hukushima, K., Sakai, Y.: An irreversible Markov-chain Monte Carlo method with skew detailed balance conditions. *J. Phys.* **473**, 012012 (2013)
28. Hwang, C.R., Hwang-Ma, S.Y., Sheu, S.J.: Accelerating Gaussian diffusions. *Ann. Appl. Probab.* **3**(3), 897–913 (1993)
29. Hwang, C.R., Hwang-Ma, S.Y., Sheu, S.J., et al.: Accelerating diffusions. *Ann. Appl. Probab.* **15**(2), 1433–1444 (2005)

30. Hwang, C.-R., Normand, R., Wu, S.-J.: Variance reduction for diffusions. [arXiv:1406.4657](https://arxiv.org/abs/1406.4657) (2014)
31. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes, vol. 1943877. Springer, Berlin (1987)
32. Kipnis, C., Varadhan, S.R.S.: Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**(1), 1–19 (1986)
33. Komorowski, T., Landim, C., Olla, S.: Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation. Springer, Heidelberg (2012)
34. Krengel, U.: Ergodic Theorems. de Gruyter Studies in Mathematics, vol. 6. Walter de Gruyter & Co., Berlin (1985)
35. Leimkuhler, B., Reich, S.: Simulating Hamiltonian Dynamics, vol. 14. Cambridge University Press, Cambridge (2004)
36. Lelièvre, T.: Two mathematical tools to analyze metastable stochastic processes. In: Cangiani, A., Davidchack, R.L., Georgoulis, E., Gorban, A.N., Levesley, J., Tretyakov, M.V. (eds.) Numerical Mathematics and Advanced Applications 2011, pp. 791–810. Springer, New York (2013)
37. Lelièvre, T., Nier, F., Pavliotis, G.A.: Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.* **152**(2), 237–274 (2013)
38. Lelièvre, T., Stoltz, G., Rousset, M.: Free Energy Computations: A Mathematical Perspective. World Scientific, Singapore (2010)
39. Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer, New York (2008)
40. Lorenzi, L., Bertoldi, M.: Analytical Methods for Markov Semigroups. CRC Press, New York (2006)
41. Ma, Y.-A., Chen, T., Fox, E.: A complete recipe for stochastic gradient mcmc. In: Advances in Neural Information Processing Systems, pp. 2899–2907 (2015)
42. MacKay, D.J.C.: Information Theory, Inference and Learning Algorithms. Cambridge University Press, Cambridge (2003)
43. Majda, A.J., Kramer, P.R.: Simplified models for turbulent diffusion: theory, numerical modelling, and physical phenomena. *Phys. Rep.* **314**(4), 237–574 (1999)
44. Majda, A.J., McLaughlin, R.M.: The effect of mean flows on enhanced diffusivity in transport by incompressible periodic velocity fields. *Stud. Appl. Math.* **89**(3), 245–279 (1993)
45. Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O.: A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.* **34**(3), A1460–A1487 (2012)
46. Mattingly, J.C., Stuart, A.M., Higham, D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Processes Appl.* **101**(2), 185–232 (2002)
47. Mattingly, J.C., Stuart, A.M., Tretyakov, M.V.: Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.* **48**(2), 552–577 (2010)
48. Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.* **25**(3), 518–548 (1993)
49. Meyn, S. P., Tweedie, R. L.: A survey of Foster-Lyapunov techniques for general state space Markov processes. In: Proceedings of the Workshop on Stochastic Stability and Stochastic Stabilization, Metz, France. Citeseer (1993)
50. Mira, A.: Ordering and improving the performance of Monte Carlo Markov chains. *Stat. Sci.* **16**(4), 340–350 (2001)
51. Neal, R.M.: Sampling from multimodal distributions using tempered transitions. *Stat. Comput.* **6**(4), 353–366 (1996)
52. Neal, R. M.: Improving asymptotic variance of MCMC estimators: nonreversible chains are better. [arXiv:math/0407281](https://arxiv.org/abs/math/0407281) (2004)
53. Neal, R.M.: MCMC Using Hamiltonian Dynamics. Handbook of Markov Chain Monte Carlo, vol. 54. CRC Press, Boca Raton (2010)
54. Ohzeki, M., Ichiki, A.: Simple implementation of Langevin dynamics without detailed balance condition. [arXiv:1307.0434](https://arxiv.org/abs/1307.0434) (2013)
55. Pardoux, E., Veretennikov, A.Y.: On the Poisson equation and diffusion approximation I. *Ann. Probab.* **29**(3), 1061–1085 (2001)
56. Pavliotis, G.: Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations, vol. 60. Springer, New York (2014)
57. Pavliotis, G. A.: Homogenization Theory for Advection—Diffusion Equations with Mean Flow, Ph.D Thesis. Rensselaer Polytechnic Institute, Troy, NY (2002)
58. Pavliotis, G.A.: Asymptotic analysis of the Green-Kubo formula. *IMA J. Appl. Math.* **75**, 951–967 (2010)
59. Peskun, P.H.: Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60**(3), 607–612 (1973)
60. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, vol. 293. Springer, New York (1999)
61. Rey-Bellet, L., Spiliopoulos, K.: Irreversible Langevin samplers and variance reduction: a large deviation approach. [arXiv:1404.0105](https://arxiv.org/abs/1404.0105) (2014)

62. Rey-Bellet, L., Spiliopoulos, K.: Variance reduction for irreversible Langevin samplers and diffusion on graphs. [arXiv:1410.0255](https://arxiv.org/abs/1410.0255) (2014)
63. Robert, C., Casella, G.: Monte Carlo Statistical Methods. Springer, New York (2013)
64. Roberts, G.O., Stramer, O.: Langevin diffusions and Metropolis-Hastings algorithms. *Methodol. Comput. Appl. Probab.* **4**(4), 337–357 (2002)
65. Roberts, G.O., Tweedie, R.L.: Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli* **2**(4), 341–363 (1996)
66. Schütte, C., Sarich, M.: Metastability and Markov State Models in molecular Dynamics: Modeling, Analysis, Algorithmic Approaches. American Mathematical Society, Providence (2013)
67. Strang, G.: On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.* **5**(3), 506–517 (1968)
68. Sun, Y., Schmidhuber, J., Faustino, J.G.: Improving the asymptotic performance of Markov chain Monte-Carlo by inserting vortices. In: Lafferty, J.D., Williams, C.K.I., Shawe-Taylor, J., Zemel, R.S., Culotta, A. (eds.) *Advances in Neural Information Processing Systems 23*, pp. 2235–2243. Curran Associates, Inc., New York (2010)
69. Suwa, H., Todo, S.: General construction of irreversible kernel in Markov Chain Monte Carlo. [arXiv:1207.0258](https://arxiv.org/abs/1207.0258) (2012)
70. Talay, D., Tubaro, L.: Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8**(4), 483–509 (1990)
71. Turitsyn, K.S., Chertkov, M., Vucelja, M.: Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D* **240**(4), 410–414 (2011)
72. Wu, S.J., Hwang, C.R., Chu, M.T.: Attaining the optimal Gaussian diffusion acceleration. *J. Stat. Phys.* **155**(3), 571–590 (2014)