J Stat Phys (2017) 169:1098–1131
DOI 10.1007/s10955-017-1906-8

Using Perturbed Underdamped Langevin Dynamics to Efficiently Sample from Probability Distributions

A. B. Duncan[1] · **N. Nüsken**[2] · **G. A. Pavliotis**[2]

Received: 30 April 2017 / Accepted: 25 October 2017 / Published online: 2 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract In this paper we introduce and analyse Langevin samplers that consist of perturbations of the standard underdamped Langevin dynamics. The perturbed dynamics is such that its invariant measure is the same as that of the unperturbed dynamics. We show that appropriate choices of the perturbations can lead to samplers that have improved properties, at least in terms of reducing the asymptotic variance. We present a detailed analysis of the new Langevin sampler for Gaussian target distributions. Our theoretical results are supported by numerical experiments with non-Gaussian target measures.

1 Introduction and Motivation

Sampling from probability measures in high-dimensional spaces is a problem that appears frequently in applications, e.g. in computational statistical mechanics and in Bayesian statistics. In particular, we are faced with the problem of computing expectations with respect to a probability measure π on $\mathbb{R}^d$, i.e. we wish to evaluate integrals of the form:

$$\pi(f) := \int_{\mathbb{R}^d} f(x)\pi(\mathrm{d}x). \tag{1}$$

As is typical in many applications, particularly in molecular dynamics and Bayesian inference, the density (for convenience denoted by the same symbol π) is known only up to a

✉ N. Nüsken
n.nusken14@imperial.ac.uk

A. B. Duncan
Andrew.Duncan@sussex.ac.uk

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

[1] School of Mathematical and Physical Sciences, University of Sussex, Falmer, Brighton BN1 9RH, UK

[2] Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, England, UK

normalization constant; furthermore, the dimension of the underlying space is quite often large enough to render deterministic quadrature schemes computationally infeasible.

A standard approach to approximating such integrals is Markov Chain Monte Carlo (MCMC) techniques [19,32,52], where a Markov process $(X_t)_{t\geq 0}$ is constructed which is ergodic with respect to the probability measure π. Then, defining the long-time average

$$\pi_T(f) := \frac{1}{T}\int_0^T f(X_s)\mathrm{d}s \tag{2}$$

for $f \in L^1(\pi)$, the ergodic theorem guarantees almost sure convergence of the long-time average $\pi_T(f)$ to $\pi(f)$.

There are infinitely many Markov, and, for the purposes of this paper diffusion, processes that can be constructed in such a way that they are ergodic with respect to the target distribution. A natural question is then how to choose the ergodic diffusion process $(X_t)_{t\geq 0}$. Naturally the choice should be dictated by the requirement that the computational cost of (approximately) calculating (1) is minimized. A standard example is given by the *overdamped Langevin dynamics* defined to be the unique (strong) solution $(X_t)_{t\geq 0}$ of the following stochastic differential equation (SDE):

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \tag{3}$$

where $V = -\log\pi$ is the potential associated with the smooth positive density π. Under appropriate assumptions on V, i.e. on the measure $\pi(\mathrm{d}x)$, the process $(X_t)_{t\geq 0}$ is ergodic and in fact reversible with respect to the target distribution.

Another well-known example is the *underdamped Langevin dynamics* given by $(X_t)_{t\geq 0} = (q_t, p_t)_{t\geq 0}$ defined on the extended space (phase space) $\mathbb{R}^d \times \mathbb{R}^d$ by the following pair of coupled SDEs:

$$\mathrm{d}q_t = M^{-1}p_t\mathrm{d}t, \quad \mathrm{d}p_t = -\nabla V(q_t)\mathrm{d}t - \Gamma M^{-1}p_t\mathrm{d}t + \sqrt{2\Gamma}\mathrm{d}W_t, \tag{4}$$

where the mass and friction tensors M and Γ are assumed to be symmetric positive definite matrices. It is well-known [36,46] that $(q_t, p_t)_{t\geq 0}$ is ergodic with respect to the measure $\widehat{\pi} := \pi \otimes \mathcal{N}(0, M)$, having density with respect to the Lebesgue measure on $\mathbb{R}^{2d}$ given by

$$\widehat{\pi}(q, p) = \frac{1}{\widehat{Z}}\exp\left(-V(q) - \frac{1}{2}p\cdot M^{-1}p\right), \tag{5}$$

where $\widehat{Z}$ is a normalization constant. Note that $\widehat{\pi}$ has marginal π with respect to p and thus for functions $f \in L^1(\pi)$, we have that $\frac{1}{t}\int_0^t f(q_t)\,\mathrm{d}t \to \pi(f)$ almost surely. Notice also that the dynamics restricted to the q-variables is no longer Markovian. The p-variables can thus be interpreted as giving some instantaneous memory to the system, facilitating efficient exploration of the state space. Higher order Markovian models, based on a finite dimensional (Markovian) approximation of the generalized Langevin equation can also be used [12].

As there is a lot of freedom in choosing the dynamics in (2), see the discussion in Sect. 2, it is desirable to choose the diffusion process $(X_t)_{t\geq 0}$ in such a way that $\pi_T(f)$ can provide a good estimation of $\pi(f)$. The performance of the estimator (2) can be quantified in various manners. The ultimate goal, of course, is to choose the dynamics as well as the numerical discretization in such a way that the computational cost of the longtime-average estimator is minimized, for a given tolerance. The minimization of the computational cost consists of three steps: bias correction, variance reduction and choice of an appropriate discretization scheme. For the latter step see Sect. 5 and [14, Sect. 6].

Under appropriate conditions on the potential V it can be shown that both (3) and (4) converge to equilibrium exponentially fast, e.g. in relative entropy. One performance objective would then be to choose the process $(X_t)_{t\geq 0}$ so that this rate of convergence is maximised. Conditions on the potential V which guarantee exponential convergence to equilibrium, both in $L^2(\pi)$ and in relative entropy can be found in [7,39,54]. In the case when the target measure π is Gaussian, both the overdamped (3) and the underdamped (4) dynamics become generalized Ornstein–Uhlenbeck processes. For such processes the entire spectrum of the generator—or, equivalently, the Fokker–Planck operator—can be computed analytically and, in particular, an explicit formula for the L^2-spectral gap can be obtained [38,43,44]. A detailed analysis of the convergence to equilibrium in relative entropy for stochastic differential equations with linear drift, i.e. generalized Ornstein–Uhlenbeck processes, has been carried out in [1,2].

In addition to speeding up convergence to equilibrium, i.e. reducing the bias of the estimator (2), one is naturally also interested in reducing the asymptotic variance. Under appropriate conditions on the target measure π and the observable f, the estimator $\pi_T(f)$ satisfies a central limit theorem (CLT) [31], that is,

$$\frac{1}{\sqrt{T}}\left(\pi_T(f)-\pi(f)\right) \xrightarrow[T\to\infty]{d} \mathcal{N}\left(0, 2\,\sigma_f^2\right),$$

where $\sigma_f^2 < \infty$ is the *asymptotic variance* of the estimator $\pi_T(f)$. The asymptotic variance characterises the magnitude of fluctuations of $\pi_T(f)$ around $\pi(f)$. Consequently, another natural objective is to choose the process $(X_t)_{t\geq 0}$ such that σ_f^2 is as small as possible. It is well known that the asymptotic variance can be expressed in terms of the solution to an appropriate Poisson equation for the generator of the dynamics [31]

$$-\mathcal{L}\phi = f - \pi(f), \quad \sigma_f^2 = \int_{\mathbb{R}^d} \phi(-\mathcal{L}\phi)\,\pi(\mathrm{d}x). \tag{6}$$

Techniques from the theory of partial differential equations can then be used in order to study the problem of minimizing the asymptotic variance. This is the approach that was taken in [14], see also [23], and it will also be used in this paper.

Other measures of performance have also been considered. For example, in [50,51], performance of the estimator is quantified in terms of the rate functional of the ensemble measure $\frac{1}{t}\int_0^t \delta_{X(t)}(dx)$. See also [28] for a study of the nonasymptotic behaviour of MCMC techniques, including the case of overdamped Langevin dynamics.

Similar analyses have been carried out for various modifications of (3). Of particular interest to us are the *Riemannian manifold MCMC* [18] (see the discussion in Sect. 2) and the *nonreversible Langevin samplers* [20,21]. As a particular example of the general framework that was introduced in [18], we mention the preconditioned overdamped Langevin dynamics $\mathrm{d}X_t = -P\nabla V(X_t)\,\mathrm{d}t + \sqrt{2P}\,\mathrm{d}W_t$, that was presented in [4]. There, the long-time behaviour of as well as the asymptotic variance of the corresponding estimator $\pi_T(f)$ are studied and applied to equilibrium sampling in molecular dynamics. A variant of the standard underdamped Langevin dynamics that can be thought of as a form of preconditioning and that has been used by practitioners is the *mass-tensor molecular dynamics* [6].

The nonreversible overdamped Langevin dynamics

$$\mathrm{d}X_t = -\left(\nabla V(X_t) - \gamma(X_t)\right)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{7}$$

where the vector field γ satisfies $\nabla \cdot (\pi\gamma) = 0$ is ergodic (but not reversible) with respect to the target measure π for all choices of the divergence-free vector field γ. The asymptotic behaviour of this process was considered for Gaussian diffusions in [20], where the rate of convergence of the covariance to equilibrium was quantified in terms of the choice of γ. This work was extended to the case of non-Gaussian target densities, and consequently for nonlinear SDEs of the form (7) in [21]. The problem of constructing the optimal nonreversible perturbation, in terms of the $L^2(\pi)$ spectral gap for Gaussian target densities was studied in [34] see also [55]. Optimal nonreversible perturbations with respect to miniziming the asymptotic variance were studied in [14,23]. In all these works it was shown that, in theory [i.e. without taking into account the computational cost of the discretization of the dynamics (7)], the nonreversible Langevin sampler (7) is never worse that its reversible counterpart (3), both in terms of converging faster to the target distribution as well as in terms of having a lower asymptotic variance. It should be emphasized that the two optimality criteria, maximizing the spectral gap and minimizing the asymptotic variance, lead to different choices for the nonreversible drift $\gamma(x)$.

The goal of this paper is to extend the analysis presented in [14,34] by introducing the following modification of the standard underdamped Langevin dynamics:

$$\begin{aligned} \mathrm{d}q_t &= M^{-1} p_t \mathrm{d}t - \mu J_1 \nabla V(q_t)\mathrm{d}t, \\ \mathrm{d}p_t &= -\nabla V(q_t)\mathrm{d}t - \nu J_2 M^{-1} p_t \mathrm{d}t - \Gamma M^{-1} p_t \mathrm{d}t + \sqrt{2\Gamma}\mathrm{d}W_t, \end{aligned} \tag{8}$$

where $M, \Gamma \in \mathbb{R}^{d\times d}$ are constant strictly positive definite matrices, μ and ν are scalar constants and $J_1, J_2 \in \mathbb{R}^{d\times d}$ are constant skew-symmetric matrices. As demonstrated in Sect. 2, the process defined by (8) will be ergodic with respect to the Gibbs measure $\widehat{\pi}$ defined in (5).

Our objective is to investigate the use of these dynamics for computing ergodic averages of the form (2). To this end, we study the long time behaviour of (8) and, using hypocoercivity techniques, prove that the process converges exponentially fast to equilibrium. This perturbed underdamped Langevin process introduces a number of parameters in addition to the mass and friction tensors which must be tuned to ensure that the process is an efficient sampler. For Gaussian target densities, we derive estimates for the spectral gap and the asymptotic variance, valid in certain parameter regimes. Moreover, for certain classes of observables, we are able to identify the choices of parameters which lead to the optimal performance in terms of asymptotic variance. While these results are valid for Gaussian target densities, we advocate these particular parameter choices also for more complex target densities. To demonstrate their efficacy, we perform a number of numerical experiments on more complex, multimodal distributions. In particular, we use the Langevin sampler (8) in order to study the problem of diffusion bridge sampling.

The rest of the paper is organized as follows. In Sect. 2 we present some background material on Langevin dynamics, we construct general classes of Langevin samplers and we introduce criteria for assessing the performance of the samplers. In Sect. 3 we study qualitative properties of the perturbed underdamped Langevin dynamics (8) including exponentially fast convergence to equilibrium and the overdamped limit. In Sect. 4 we study in detail the performance of the Langevin sampler (8) for the case of Gaussian target distributions. In Sect. 5 we introduce a numerical scheme for simulating the perturbed dynamics (8) and we present numerical experiments on the implementation of the proposed samplers for the problem of diffusion bridge sampling. Section 6 is reserved for conclusions and suggestions for further work.

2 Construction of General Langevin Samplers

2.1 Background and Preliminaries

In this section we consider estimators of the form (2) where $(X_t)_{t\geq 0}$ is a diffusion process given by the solution of the following Itô SDE:

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \sqrt{2}b(X_t)\,\mathrm{d}W_t, \tag{9}$$

with drift coefficient $a : \mathbb{R}^d \to \mathbb{R}^d$ and diffusion coefficient $b : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ both having smooth components, and where $(W_t)_{t\geq 0}$ is a standard $\mathbb{R}^m$–valued Brownian motion. Associated with (9) is the infinitesimal generator $\mathcal{L}$ formally given by

$$\mathcal{L}f = a\cdot\nabla f + \Sigma : \nabla\nabla f, \quad f \in C_c^2(\mathbb{R}^d) \tag{10}$$

where $\Sigma = bb^\top$, $\nabla\nabla f$ denotes the Hessian of the function f and : denotes the Frobenius inner product. In general, Σ is nonnegative definite, and could possibly be degenerate. In particular, the infinitesimal generator (10) need not be uniformly elliptic. To ensure that the corresponding semigroup exhibits sufficient smoothing behaviour, we shall require that the process (9) is hypoelliptic in the sense of Hörmander. If this condition holds, then irreducibility of the process $(X_t)_{t\geq 0}$ will be an immediate consequence of the existence of a strictly positive invariant distribution $\pi(x)\mathrm{d}x$, see [30].

Suppose that $(X_t)_{t\geq 0}$ is nonexplosive. It follows from the hypoellipticity assumption that the process $(X_t)_{t\geq 0}$ possesses a smooth transition density $p(t, x, y)$ which is defined for all $t \geq 0$ and $x, y \in \mathbb{R}^d$, [5, Theorem VII.5.6]. The associated strongly continuous Markov semigroup $(P_t)_{t\geq 0}$ is defined by $P_t f(x) = \int_{\mathbb{R}^d} p(t, x, y) f(y)\,\mathrm{d}y$. Suppose that $(P_t)_{t\geq 0}$ is invariant with respect to the target measure π, i.e. $\int_{\mathbb{R}^d} P_t f(x)\pi(\mathrm{d}x) = \int_{\mathbb{R}^d} f(x)\pi(\mathrm{d}x)$ for $t \geq 0$ and all bounded continuous functions f. Then $(P_t)_{t\geq 0}$ can be extended to a positivity preserving contraction semigroup on $L^2(\pi)$ which is strongly continuous. Moreover, the infinitesimal generator corresponding to $(P_t)_{t\geq 0}$ is given by an extension of $(\mathcal{L}, C_c^2(\mathbb{R}^d))$, also denoted by $\mathcal{L}$.

Due to hypoellipticity and invariance with respect to $(P_t)_{t\geq 0}$, the probability measure π on $\mathbb{R}^d$ has a smooth density with respect to the Lebesgue measure. If this density is strictly positive, it follows that π is necessarily the unique invariant distribution. Slightly abusing the notation, we will denote both the measure and its density by π. Furthermore, we will denote by $L^2(\pi)$ be the Hilbert space of π-square integrable functions equipped with inner product $\langle\cdot,\cdot\rangle_{L^2(\pi)}$ and norm $\|\cdot\|_{L^2(\pi)}$.

2.2 A General Characterisation of Ergodic Diffusions

A natural question is what conditions on the coefficients a and b of (9) are required to ensure that $(X_t)_{t\geq 0}$ is invariant with respect to the distribution $\pi(x)\,\mathrm{d}x$. The following result provides a necessary and sufficient condition for a diffusion process to be invariant with respect to a given target distribution.

Theorem 1 *Consider a diffusion process $(X_t)_{t\geq 0}$ on $\mathbb{R}^d$ defined by the unique, nonexplosive solution to the Itô SDE (9) with drift $a \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ and diffusion coefficient $b \in C^1(\mathbb{R}^d; \mathbb{R}^{d\times m})$. Then $(X_t)_{t\geq 0}$ is invariant with respect to π if and only if*

$$a = \Sigma\nabla\log\pi + \nabla\cdot\Sigma + \gamma, \tag{11}$$

where $\Sigma = bb^\top$ *and* $\gamma : \mathbb{R}^D \to \mathbb{R}^D$ *is a continuously differentiable vector field satisfying*

$$\nabla \cdot (\pi\gamma) = 0. \tag{12}$$

If additionally $\gamma \in L^1(\pi)$*, then there exists a skew-symmetric matrix function* $C : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *such that* $\gamma = \frac{1}{\pi}\nabla \cdot (\pi C)$ *. In this case the infinitesimal generator can be written as an* $L^2(\pi)$*-extension of*

$$\mathcal{L}f = \frac{1}{\pi}\nabla \cdot ((\Sigma + C)\pi\nabla f), \quad f \in C_c^2(\mathbb{R}^d).$$

The proof of the first part of this result can be found in [46, Chap. 4]; similar versions of this characterisation can be found in [54] and [21]. For the existence of the skew-symmetric matrix C see, e.g., [16, Sec.4, Prop. 1]. See also [37].

Remark 1 If (11) holds and $\mathcal{L}$ is hypoelliptic it follows immediately that $(X_t)_{t\geq 0}$ is ergodic with unique invariant distribution $\pi(x)\,\mathrm{d}x$ (see [30]).

More generally, we can consider Itô diffusions in an extended phase space:

$$\mathrm{d}Z_t = b(Z_t)\,\mathrm{d}t + \sqrt{2}\sigma(Z_t)\,\mathrm{d}W_t, \tag{13}$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}^N$, $N \geq d$. This is a Markov process with generator

$$\mathcal{L} = b(z)\cdot\nabla_z + \Sigma(z) : \nabla_z\nabla_z, \tag{14}$$

where $\Sigma(z) = \left(\sigma\sigma^T\right)(z)$. We will consider dynamics $(Z_t)_{t\geq 0}$ that is ergodic with respect to $\pi_z(z)\,\mathrm{d}z$ such that $\int_{\mathbb{R}^m}\pi_z(x,\,y)\,\mathrm{d}y = \pi(x)$, where $z = (x,\,y),\ x \in \mathbb{R}^d,\ y \in \mathbb{R}^m,\ d+m = N$.

There are various well-known choices of dynamics which are invariant (and indeed ergodic) with respect to the target distribution $\pi(x)\mathrm{d}x$.

1. Choosing $b = I$ and $\gamma = 0$ we immediately recover the overdamped Langevin dynamics (3).
2. Choosing $b = I$, and $\gamma \neq 0$ such that (12) holds gives rise to the nonreversible overdamped equation defined by (7). As it satisfies the conditions of Theorem 1, it is ergodic with respect to π. In particular choosing $\gamma(x) = J\nabla V(x)$ for a constant skew-symmetric matrix J we obtain

$$\mathrm{d}X_t = -(I+J)\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{15}$$

   which has been studied in previous works.
3. Given a target density $\pi > 0$ on $\mathbb{R}^d$, if we consider the augmented target density $\widehat{\pi}$ on $\mathbb{R}^{2d}$ given in (5), then choosing

$$\gamma((q,\,p)) = \begin{pmatrix} M^{-1}p \\ -\nabla V(q)\end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix}\mathbf{0}\\ \sqrt{\Gamma}\end{pmatrix} \in \mathbb{R}^{2d\times d}, \tag{16}$$

   where M and Γ are positive definite symmetric matrices, the conditions of Theorem 1 are satisfied for the target density $\widehat{\pi}$. The resulting dynamics $(q_t,\,p_t)_{t\geq 0}$ is determined by the underdamped Langevin equation (4). It is straightforward to verify that the generator is hypoelliptic, [35, Sec 2.2.3.1], and thus $(q_t,\,p_t)_{t\geq 0}$ is ergodic.
4. More generally, consider the augmented target density $\widehat{\pi}$ on $\mathbb{R}^{2d}$ as above, and choose

$$\gamma((q,\,p)) = \begin{pmatrix} M^{-1}p - \mu J_1\nabla V(q) \\ -\nabla V(q) - \nu J_2 M^{-1}p\end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix}\mathbf{0}\\ \sqrt{\Gamma}\end{pmatrix} \in \mathbb{R}^{2d\times d}, \tag{17}$$

where μ and ν are scalar constants and $J_1, J_2 \in \mathbb{R}^{d\times d}$ are constant skew-symmetric matrices. With this choice we recover the perturbed Langevin dynamics (8). It is straightforward to check that (17) satisfies the invariance condition (12), and thus Theorem 1 guarantees that (8) is invariant with respect to $\widehat{\pi}$.

5. In a similar fashion, one can introduce an augmented target density on $\mathbb{R}^{(m+2)d}$, with

$$\widehat{\widehat{\pi}}(q, p, u_1, \ldots, u_m) \propto e^{-\left(\frac{|p|^2}{2}+\frac{|u_1|^2+\ldots+|u_m|^2}{2}+V(q)\right)},$$

where $p, q, u_i \in \mathbb{R}^d$, for $i = 1, \ldots, m$. Clearly $\int_{\mathbb{R}^d\times\mathbb{R}^{md}} \widehat{\widehat{\pi}}(q, p, u_1, \ldots, u_m)\,\mathrm{d}p\,\mathrm{d}u_1 \ldots \mathrm{d}u_m = \pi(q)$. We now define $\gamma : \mathbb{R}^{(m+2)d} \to \mathbb{R}^{(m+2)d}$ by

$$\gamma(q, p, u_1, \ldots, u_m) = \begin{pmatrix} p & -\nabla_q V(q) + \sum_{j=1}^{m} \lambda_j u_j & -\lambda_1 p & \cdots & -\lambda_m p \end{pmatrix}^T$$

and $b : \mathbb{R}^{(m+2)d} \to \mathbb{R}^{(m+2)d\times(m+2)d}$ by

$$b(q, p, u_1, \ldots, u_m) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sqrt{\alpha_1} I_{d\times d} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\alpha_2} I_{d\times d} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \sqrt{\alpha_m} I_{d\times d} \end{pmatrix},$$

where $\lambda_i \in \mathbb{R}$ and $\alpha_i > 0$, for $i = 1, \ldots, m$. The resulting process (9) is given by

$$\begin{aligned} \mathrm{d}q_t &= p_t\,\mathrm{d}t, \quad \mathrm{d}p_t = -\nabla_q V(q_t)\,\mathrm{d}t + \sum_{j=1}^{d} \lambda_j u^j(t)\,\mathrm{d}t, \\ \mathrm{d}u_t^1 &= -\lambda_1 p_t\,\mathrm{d}t - \alpha_1 u_t^1\,\mathrm{d}t + \sqrt{2\alpha_1}\,\mathrm{d}W_t^1, \\ &\vdots \\ \mathrm{d}u_t^m &= -\lambda_m p_t\,\mathrm{d}t - \alpha_m u_t^m\,\mathrm{d}t + \sqrt{2\alpha_m}\,\mathrm{d}W_t^m, \end{aligned} \tag{18}$$

where $(W_t^j)_{t\geq 0, j=1,\ldots,m}$ are independent $\mathbb{R}^d$–valued Brownian motions. This process is ergodic with unique invariant distribution $\widehat{\widehat{\pi}}$, and under appropriate conditions on V, converges exponentially fast to equilibrium in relative entropy [42]. Equation (18) is a Markovian representation of a generalised Langevin equation of the form

$$\mathrm{d}q_t = p_t\,\mathrm{d}t, \quad \mathrm{d}p_t = -\nabla_q V(q_t)\,\mathrm{d}t - \int_0^t F(t-s)p_s\,\mathrm{d}s + N(t),$$

where $N(t)$ is a mean-zero stationary Gaussian process with autocorrelation function $F(t)$, i.e. $\mathbb{E}\,[N(t)\otimes N(s)] = F(t-s)I_{d\times d}$ and $F(t) = \sum_{i=1}^{m} \lambda_i^2 e^{-\alpha_i|t|}$.

6. Let $\widetilde{\pi}(z) \propto \exp(-\Phi(z))$ be a positive density on $\mathbb{R}^N$ where $N > d$ such that $\pi(x) = \int_{\mathbb{R}^{N-d}} \widetilde{\pi}(x, z)\,\mathrm{d}z$, where $(x, y) \in \mathbb{R}^d \times \mathbb{R}^{N-d}$. Then choosing $b = I_{D\times D}$ and $\gamma = 0$ we obtain the dynamics

$$\mathrm{d}X_t = -\nabla_x \Phi(X_t, Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^1, \quad \mathrm{d}Y_t = -\nabla_y \Phi(X_t, Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^2,$$

then $(X_t, Y_t)_{t\geq 0}$ is immediately ergodic with respect to $\widetilde{\pi}$.

2.3 Comparison Criteria

For a fixed observable f, a natural measure of accuracy of the estimator $\pi_T(f) = t^{-1}\int_0^t f(X_s)\,\mathrm{d}s$ is the *mean square error* (MSE) defined by

$$MSE(f,T) := \mathbb{E}_x\left|\pi_T(f) - \pi(f)\right|^2, \tag{19}$$

where $\mathbb{E}_x$ denotes the expectation conditioned on the process $(X_t)_{t\geq 0}$ starting at x. It is instructive to introduce the decomposition $MSE(f,T) = \mu^2(f,T) + \sigma^2(f,T)$, where

$$\begin{aligned}\mu(f,T) &= \left|\mathbb{E}_x[\pi_T(f)] - \pi(f)\right| \quad \text{and} \quad \sigma^2(f,T) = \mathbb{E}_x\left|\pi_T(f) - \mathbb{E}_x\pi_T(f)\right|^2 \\ &= \mathrm{Var}[\pi_T(f)].\end{aligned} \tag{20}$$

Here $\mu(f,T)$ measures the bias of the estimator $\pi_T(f)$ and $\sigma^2(f,T)$ measures the variance of fluctuations of $\pi_T(f)$ around the mean.

The speed of convergence to equilibrium of the process $(X_t)_{t\geq 0}$ will control both the bias term $\mu(f,T)$ and the variance $\sigma^2(f,T)$. To make this claim more precise, suppose that the semigroup $(P_t)_{t\geq 0}$ associated with $(X_t)_{t\geq 0}$ decays exponentially fast in $L^2(\pi)$, i.e. there exist constants $\lambda > 0$ and $C \geq 1$ such that

$$\left\|P_t g - \pi(g)\right\|_{L^2(\pi)} \leq Ce^{-\lambda t}\left\|g - \pi(g)\right\|_{L^2(\pi)}, \quad g \in L^2(\pi). \tag{21}$$

Remark 2 If (21) holds with $C = 1$, this estimate is equivalent to $-\mathcal{L}$ having a spectral gap in $L^2(\pi)$. Allowing for a constant $C > 1$ is essential for our purposes though in order to treat nonreversible and degenerate diffusion processes by the theory of *hypocoercivity* as outlined in [54].

The following lemma characterises the decay of the bias $\mu(f,T)$ as $T \to \infty$ in terms of λ and C. The proof can be found in [41].

Lemma 1 *Let $(X_t)_{t\geq 0}$ be the unique, non-explosive solution of (9), such that $X_0 \sim \pi_0 \ll \pi$ and $\frac{d\pi_0}{d\pi} \in L^2(\pi)$, where $\frac{d\pi_0}{d\pi}$ denotes the Radon-Nikodym derivative of π_0 with respect to π. Suppose that the process is ergodic with respect to π such that the Markov semigroup $(P_t)_{t\geq 0}$ satisfies (21). Then for $f \in L^\infty(\pi)$,*

$$\mu(f,T) \leq \frac{C}{\lambda T}\left(1 - e^{-\lambda T}\right)\|f\|_{L^\infty}\mathrm{Var}_\pi\left[\frac{d\pi_0}{d\pi}\right]^{\frac{1}{2}}.$$

The study of the long time behaviour of the variance $\sigma^2(f,T)$ involves deriving a central limit theorem for the additive functional $\int_0^t f(X_t) - \pi(f)\,\mathrm{d}t$. As discussed in [13], we reduce this problem to proving well-posedness of the Poisson equation

$$-\mathcal{L}\chi = f - \pi(f), \quad \pi(\chi) = 0. \tag{22}$$

The only complications in this approach arise from the fact that the generator $\mathcal{L}$ need not be symmetric in $L^2(\pi)$ nor uniformly elliptic. The following result summarises conditions for the well-posedness of the Poisson equation and it also provides with us with a formula for the asymptotic variance. Again, the proof can be found in [41].

Lemma 2 *Let $(X_t)_{t\geq 0}$ be the unique, non-explosive solution of (9) with smooth drift and diffusion coefficients, such that the corresponding infinitesimal generator is hypoelliptic. Syppose that $(X_t)_{t\geq 0}$ is ergodic with respect to π and moreover, $(P_t)_{t\geq 0}$ decays exponentially*

fast in $L^2(\pi)$ *as in* (21)*. Then for all* $f \in L^2(\pi)$*, there exists a unique mean zero solution* χ *to the Poisson equation* (22)*. If* $X_0 \sim \pi$*, then for all* $f \in C^\infty(\mathbb{R}^d) \cap L^2(\pi)$

$$\sqrt{T}\left(\pi_T(f) - \pi(f)\right) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, 2\sigma_f^2), \tag{23}$$

where σ_f^2 *is the asymptotic variance defined by*

$$\sigma_f^2 = \langle \chi, (-\mathcal{L})\chi \rangle_{L^2(\pi)} = \langle \nabla\chi, \Sigma\nabla\chi \rangle_{L^2(\pi)}. \tag{24}$$

Moreover, if $X_0 \sim \pi_0$ *where* $\pi_0 \ll \pi$ *and* $\frac{d\pi_0}{d\pi} \in L^2(\pi)$ *then* (23) *holds for all* $f \in C^\infty(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$.[1]

Clearly, observables that only differ by a constant have the same asymptotic variance. In the sequel, we will hence restrict our attention to observables $f \in L^2(\pi)$ satisfying $\pi(f) = 0$, simplifying expressions (22) and (23). The corresponding subspace of $L^2(\pi)$ will be denoted by $L_0^2(\pi)$. If the exponential decay estimate (21) is satisfied, then Lemma 2 shows that $-\mathcal{L}$ is invertible on $L_0^2(\pi)$, so we can express the asymptoptic variance as

$$\sigma_f^2 = \langle f, (-\mathcal{L})^{-1} f \rangle_{L^2(\pi)}, \quad f \in L_0^2(\pi). \tag{25}$$

We note that the constants C and λ appearing in the exponential decay estimate (21) also control the speed of convergence of $\sigma^2(f, T)$ to zero. Indeed, it is straightforward to show that if (21) is satisfied, then the solution χ of (22) satisfies

$$\sigma_f^2 = \langle \chi, f - \pi(f) \rangle_{L^2(\pi)} \leq \frac{C}{\lambda} \|f\|_{L^2(\pi)}^2. \tag{26}$$

Lemmas 1 and 2 would suggest that choosing the coefficients Σ and γ to optimize the constants C and λ in (26) would be an effective means of improving the performance of the estimator $\pi_T(f)$, especially since the improvement in performance would be uniform over an entire class of observables. When this is possible, this is indeed the case. However, as has been observed in [20,21,34], maximising the speed of convergence to equilibrium is a delicate task. As the leading order term in $MSE(f, T)$, it is typically sufficient to focus specifically on the asymptotic variance σ_f^2 and study how the parameters of the SDE (9) can be chosen to minimise σ_f^2. This study was undertaken in [14] for processes of the form (7).

3 Perturbation of Underdamped Langevin Dynamics

The primary objective of this work is to compare the performances of the perturbed underdamped Langevin dynamics (8) and the unperturbed dynamics (4) according to the criteria outlined in Sect. 2.3 and to find suitable choices for the matrices J_1, J_2, M and Γ that improve the performance of the sampler. We begin our investigations of (8) by establishing ergodicity and exponentially fast return to equilibrium, and by studying the overdamped limit of (8). As the latter turns out to be nonreversible and therefore in principle superior to the usual overdamped limit (3), e.g. [21], this calculation provides us with further motivation to study the proposed dynamics.

For the bulk of this work, we focus on the particular case when the target measure is Gaussian, i.e. when the potential is given by $V(q) = \frac{1}{2} q^T S q$ with a symmetric and positive

[1] In fact, using the results from [8], we could consider observables in $L^2(\pi)$. However, we will not extend this point further in this paper.

definite precision matrix S (i.e. the covariance matrix is given by S^{-1}). In this case, we advocate the following conditions for the choice of parameters:

$$M = S, \quad \Gamma = \gamma S, \quad SJ_1S = J_2, \quad \mu = \nu. \tag{27}$$

Under the above choices (27), we show that the large perturbation limit $\lim_{\mu\to\infty}\sigma_f^2$ exists and is finite and we provide an explicit expression for it (see Corollary 4). From this expression, we derive an algorithm for finding optimal choices for J_1 in the case of quadratic observables (see Algorithm 2).

If the friction coefficient is not too small ($\gamma > \sqrt{2}$), and under certain mild nondegeneracy conditions, we prove that adding a small perturbation will always decrease the asymptotic variance for observables of the form $f(q) = q \cdot Kq + l \cdot q + C$:

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\sigma_f^2\Big|_{\mu=0} = 0 \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}\mu^2}\sigma_f^2\Big|_{\mu=0} < 0,$$

see Theorem 3. In fact, we conjecture that this statement is true for arbitrary observables $f \in L^2(\pi)$, but we have not been able to prove this. The dynamics (8) [used in conjunction with the conditions (27)] proves to be especially effective when the observable is antisymmetric (i.e. when it is invariant under the substitution $q \mapsto -q$) or when it has a significant antisymmetric part. In particular, in Proposition 3 we show that under certain conditions on the spectrum of J_1, for any antisymmetric observable $f \in L^2(\pi)$ it holds that $\lim_{\mu\to\infty}\sigma_f^2 = 0$.

Numerical experiments and analysis show that departing significantly from (27) in fact possibly decreases the performance of the sampler. This is in stark contrast to (7), where it is not possible to increase the asymptotic variance by *any* perturbation. For that reason, until now it seems practical to use (8) as a sampler only when a reasonable estimate of the global covariance of the target distribution is available. In the case of Bayesian inverse problems and diffusion bridge sampling, the target measure π is given with respect to a Gaussian prior. We demonstrate the effectiveness of our approach in these applications, taking the prior Gaussian covariance as S in (27).

Remark 3 In [34, Rem. 3] another modification of (4) was suggested (albeit with the simplifications $\Gamma = \gamma \cdot I$ and $M = I$):

$$\mathrm{d}q_t = (1-J)M^{-1}p_t\mathrm{d}t, \quad \mathrm{d}p_t = -(1+J)\nabla V(q_t)\mathrm{d}t - \Gamma M^{-1}p_t\mathrm{d}t + \sqrt{2\Gamma}\mathrm{d}W_t, \tag{28}$$

J again denoting an antisymmetric matrix. However, under the change of variables $p \mapsto (1+J)\tilde{p}$ the above equations transform into

$$\mathrm{d}q_t = \tilde{M}^{-1}p_t\mathrm{d}t, \quad \mathrm{d}\tilde{p}_t = -\nabla V(q_t)\mathrm{d}t - \tilde{\Gamma}\tilde{M}^{-1}\tilde{p}_t\mathrm{d}t + \sqrt{2\tilde{\Gamma}}\mathrm{d}\tilde{W}_t, \tag{29}$$

where $\tilde{M} = (1+J)^{-1}M(1-J)^{-1}$ and $\tilde{\Gamma} = (1+J)^{-1}\Gamma(1-J)^{-1}$. Since any observable f depends only on q (the p-variables are merely auxiliary), the estimator $\pi_T(f)$ as well as its associated convergence characteristics (i.e. asymptotic variance and speed of convergence to equilibrium) are invariant under this transformation. Therefore, (28) reduces to the underdamped Langevin dynamics (4) and does not represent an independent approach to sampling. Suitable choices of M and Γ will be discussed in Sect. 4.5.

3.1 Properties of Perturbed Underdamped Langevin Dynamics

In this section we study some of the properties of the perturbed underdamped dynamics (8). First, note that its generator is given by

$$\mathcal{L} = \underbrace{\underbrace{M^{-1}p\cdot\nabla_q - \nabla_q V\cdot\nabla_p}_{\mathcal{L}_{ham}} \underbrace{-\Gamma M^{-1}p\cdot\nabla_p + \Gamma : D_p^2}_{\mathcal{L}_{therm}}}_{\mathcal{L}_0} \underbrace{-\mu J_1\nabla V\cdot\nabla_q - \nu J M^{-1}p\cdot\nabla_p}_{\mathcal{L}_{pert}}, \tag{30}$$

decomposed into the perturbation $\mathcal{L}_{pert}$ and the unperturbed operator $\mathcal{L}_0$, which can be further split into the Hamiltonian part $\mathcal{L}_{ham}$ and the thermostat (Ornstein–Uhlenbeck) part $\mathcal{L}_{therm}$, see [35,36,46].

Lemma 3 *The infinitesimal generator $\mathcal{L}$ (30) is hypoelliptic.*

Proof The proof consists of verifying the conditions of Hörmander's Theorem for the generator (30) and can be found in [41]. □

An immediate corollary of this result and of Theorem 1 is that the perturbed underdamped Langevin process (8) is ergodic with unique invariant distribution $\widehat{\pi}$ given by (5).

As explained in Sect. 2.3, the exponential decay estimate (21) is crucial for our approach, as in particular it guarantees the well-posedness of the Poisson equation (22). From now on, we will therefore make the following assumption on the potential V, required to prove exponential decay in $L^2(\pi)$:

Assumption 1 Assume that the Hessian of V is *bounded* and that the target measure $\pi(\mathrm{d}q) = \frac{1}{Z}e^{-V}\mathrm{d}q$ satisfies a *Poincare inequality*, i.e. there exists a constant $\rho > 0$ such that

$$\int_{\mathbb{R}^d}\phi^2\mathrm{d}\pi \le \rho\int_{\mathbb{R}^d}|\nabla\phi|^2\mathrm{d}\pi, \quad \phi\in L_0^2(\pi)\cap H^1(\pi). \tag{31}$$

Sufficient conditions on the potential so that Poincaré's inequality holds, e.g. the Bakry-Emery criterion, are presented in [7].

Theorem 2 *Under Assumption 1 there exist constants $C \ge 1$ and $\lambda > 0$ such that the semigroup $(P_t)_{t\ge 0}$ generated by $\mathcal{L}$ satisfies exponential decay in $L^2(\pi)$ as in (21).*

Proof The proof uses the machinery of hypocoercivity developed in [54] and can be found in [41]. Using the framework of [15], we conjecture that the assumption on the boundedness of the Hessian of V can be substantially weakened and more quantitative decay estimates (in particular with respect to μ and ν) can be obtained. This approach has recently been successfully applied to equilibrium and nonequilibirum Langevin dynamics, see [27,53]. We leave this work track for future study. □

3.2 The Overdamped Limit

In this section we develop a connection between the perturbed underdamped Langevin dynamics (8) and the nonreversible overdamped Langevin dynamics (7). The analysis is very similar to the one presented in [35, Sect. 2.2.2] and we will be brief. For convenience in this section we will perform the analysis on the d-dimensional torus $\mathbb{T}^d \cong (\mathbb{R}/\mathbb{Z})^d$, i.e. we will assume $q\in\mathbb{T}^d$. Consider the following scaling of (8):

$$\mathrm{d}q_t^\epsilon = \frac{1}{\epsilon} M^{-1} p_t^\epsilon, \mathrm{d}t - \mu J_1 \nabla_q V(q_t)\mathrm{d}t, \tag{32a}$$

$$\mathrm{d}p_t^\epsilon = -\frac{1}{\epsilon}\nabla_q V(q_t^\epsilon)\mathrm{d}t - \frac{1}{\epsilon^2}\nu J_2 M^{-1} p_t^\epsilon \mathrm{d}t - \frac{1}{\epsilon^2}\Gamma M^{-1} p_t^\epsilon \mathrm{d}t + \frac{1}{\epsilon}\sqrt{2\Gamma}\mathrm{d}W_t, \tag{32b}$$

valid for the small mass/small momentum regime $M \to \epsilon^2 M$, $p_t \to \epsilon p_t$. Equivalently, those modifications can be obtained from subsituting $\Gamma \to \epsilon^{-1}\Gamma$ and $t \mapsto \epsilon^{-1}t$, and so in the limit as $\epsilon \to 0$ the dynamics (32) describes the limit of large friction with rescaled time. It turns out that as $\epsilon \to 0$, the dynamics (32) converges to the limiting SDE

$$\mathrm{d}q_t = -(\nu J_2 + \Gamma)^{-1}\nabla_q V(q_t)\mathrm{d}t - \mu J_1 \nabla_q V(q_t)\mathrm{d}t + (\nu J_2 + \Gamma)^{-1}\sqrt{2\Gamma}\mathrm{d}W_t. \tag{33}$$

The following proposition makes this statement precise.

Proposition 1 *Denote by $(q_t^\epsilon, p_t^\epsilon)$ the solution to (32) with (deterministic) initial conditions $(q_0^\epsilon, p_0^\epsilon) = (q_{init}, p_{init})$ and by q_t^0 the solution to (33) with initial condition $q_0^0 = q_{init}$. For any $T > 0$, $(q_t^\epsilon)_{0\le t\le T}$ converges to $(q_t^0)_{0\le t\le T}$ in $L^2(\Omega, C([0,T]), \mathbb{T}^d)$ as $\epsilon \to 0$, i.e. $\lim_{\epsilon\to 0}\mathbb{E}\big(\sup_{0\le t\le T}|q_t^\epsilon - q_t^0|^2\big) = 0$.*

Proof The proof follows standard arguments (see for instance [46]) and can be found in [41]. By a more refined analysis, it is also possible to get information on the rate of convergence; see, e.g. [48,49]. □

Remark 4 The overdamped limit (33) respects the invariant distribution, in the sense that it is ergodic with respect to $\pi(\mathrm{d}q) = \frac{1}{Z}e^{-V}\mathrm{d}q$.

The limiting SDE (33) is nonreversible due to the term $-\mu J_1 \nabla_q V(q_t)\mathrm{d}t$ and also because the matrix $(\nu J_2 + \Gamma)^{-1}$ is in general neither symmetric nor antisymmetric. This result, together with the fact that nonreversible perturbations of overdamped Langevin dynamics of the form (7) are by now well-known to have improved performance properties, motivates further investigation of the dynamics (8).

4 Sampling from a Gaussian Distribution

In this section we study in detail the performance of the Langevin sampler (8) for Gaussian target densities, first considering the case of unit covariance. In particular, we study the optimal choice for the parameters in the sampler, the exponential decay rate and the asymptotic variance. We then extend our results to Gaussian target densities with arbitrary covariance matrices.

4.1 Unit Covariance: Small Perturbations

In our study of the dynamics given by (8) we first consider the simple case when $V(q) = \frac{1}{2}|q|^2$, i.e. the task of sampling from a Gaussian measure with unit covariance. We will assume $M = I$, $\Gamma = \gamma I$ and $J_1 = J_2 =: J$ (so that the $q-$ and $p-$dynamics are perturbed in the same way, albeit posssibly with different strengths μ and ν). Our first result concerns the asymptotic variance for linear and quadratic observables for small perturbations of equal strength ($\mu = \nu$). For sufficiently strong damping ($\gamma > \sqrt{2}$) always leads to an improvement in asymptotic variance under the nondegeneracy conditions $[J, K] \neq 0$ and $l \notin \ker J$:

Theorem 3 *Consider the dynamics*

$$\begin{aligned} \mathrm{d}q_t &= p_t \mathrm{d}t - \mu J q_t \mathrm{d}t, \\ \mathrm{d}p_t &= -q_t \mathrm{d}t - \mu J p_t \mathrm{d}t - \gamma p_t \mathrm{d}t + \sqrt{2\gamma}\mathrm{d}W_t, \end{aligned} \tag{34}$$

with $\gamma > \sqrt{2}$ and an observable of the form $f(q) = q \cdot Kq + l \cdot q + C$, where $K \in \mathbb{R}^{d\times d}_{sym}$, $l \in \mathbb{R}^d$ and $C \in \mathbb{R}$. If at least one of the conditions $[J, K] \neq 0$ and $l \notin \ker J$ is satisfied, then the asymptotic variance of the unperturbed sampler is at a local maximum independently of K and J (and γ, as long as $\gamma > \sqrt{2}$), i.e.

$$\partial_\mu \sigma_f^2\Big|_{\mu=0} = 0 \quad \textit{and} \quad \partial_\mu^2 \sigma_f^2\Big|_{\mu=0} < 0.$$

Proof The dynamics (34) are of Ornstein–Uhlenbeck type, i.e. we can write

$$\mathrm{d}X_t = -BX_t \mathrm{d}t + \sqrt{2Q}\mathrm{d}\bar{W}_t, \quad B = \begin{pmatrix} \mu J & -I \\ I & \gamma I + \mu J \end{pmatrix}, \quad Q = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \gamma I \end{pmatrix} \tag{35}$$

with $X = (q, p)^T$, and $(\bar{W}_t)_{t\geq 0}$ denoting a standard Wiener process on $\mathbb{R}^{2d}$. The generator of (35) is then given by

$$\mathcal{L} = -Bx \cdot \nabla + \nabla^T Q \nabla. \tag{36}$$

According to Lemma 2, the asymptotic variance can be expressed as

$$\sigma_f^2 = \langle \chi, f \rangle_{L^2(\widehat{\pi})}, \quad \text{where } \chi \text{ is the solution to } \quad -\mathcal{L}\chi = f, \quad \widehat{\pi}(\chi) = 0. \tag{37}$$

By calculations similar to those in [14, Sect. 4], χ is given by $\chi(x) = x \cdot Cx + D \cdot x - \operatorname{Tr} C$, where

$$BC + CB^T = \bar{K}, \quad B^T D = \bar{l}, \tag{38}$$

using the notations

$$\bar{K} = \begin{pmatrix} K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{2d\times 2d} \quad \text{and} \quad \bar{l} = \begin{pmatrix} l \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{2d}. \tag{39}$$

The asymptotic variance is then given by

$$\sigma_f^2 = 2\operatorname{Tr}(C\bar{K}) + D \cdot \bar{l}. \tag{40}$$

Taking derivatives of 38 and solving the ensuing matrix equations, it is possible to obtain explicit expressions for $\partial_\mu C|_{\mu=0}$, $\partial_\mu^2 C|_{\mu=0}$, $\partial_\mu D|_{\mu=0}$ and $\partial_\mu^2 C|_{\mu=0}$ as detailed in [41]. We obtain

$$\begin{aligned} \partial_\mu \sigma_f^2\Big|_{\mu=0} = 0 \quad \text{and} \quad \partial_\mu^2 \sigma_f^2\Big|_{\mu=0} &= (-2\gamma^3 + 4\gamma)|Jl|^2 \\ &+ \left(\gamma - \frac{4}{\gamma^3} - \gamma^3 - \frac{1}{\gamma}\right) \cdot \left(\operatorname{Tr}(JKJK) - \operatorname{Tr}(J^2K^2)\right). \end{aligned}$$

Notice that $\operatorname{Tr}(JKJK) - \operatorname{Tr}(J^2K^2) = \frac{1}{2}\operatorname{Tr}([J, K]^2)$ and that $[J, K]$ is symmetric. It follows that $\operatorname{Tr}(JKJK) - \operatorname{Tr}(J^2K^2) \geq 0$ with equality if and only if $[J, K] = 0$. Together with $-2\gamma^3 + 4\gamma < 0$ for $\gamma > \sqrt{2}$ and $\gamma - \frac{4}{\gamma^3} - \gamma^3 - \frac{1}{\gamma} < 0$ for $\gamma > 0$, the claim follows. □

Remark 5 As we will see in Sect. 4.3, Example 1, if $[J, K] = 0$ and $l \in \ker J$, the asymptotic variance is constant as a function of μ, i.e. the perturbation has no effect.

Numerical examples show that the conditions $\gamma > \sqrt{2}$ and $\mu = \nu$ are indeed necessary for the conclusions of Theorem 3 to hold (an explanation in terms of the spectrum of the generator will be provided in Sect. 4.2). In particular, an unfortunate choice of the perturbations will actually increase the asymptotic variance of the dynamics.

Let us illustrate this by plotting the asymptotic variance as a function of the perturbation strength μ (see Fig. 1), making the choices $d = 2, l = (1, 1)^T$,

$$K = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{41}$$

The asymptotic variance has been computed according to (38) and (40). Going beyond the results of this section, the graphs give an impression of the behaviour of the asymptotic variace for large values of μ, discussed further in Sect. 4.3.

Figure 1a, b, c show the asymptotic variance associated with the quadratic observable $f(q) = q \cdot Kq$. In accordance with Theorem 3, the asymptotic variance is at a local maximum at zero perturbation in the case $\mu = \nu$ (see Fig. 1a). For increasing perturbation strength, the graph shows monotone decay for $\mu \to \infty$ (this limiting behaviour will be explored analytically in Sect. 4.3). If the condition $\mu = \nu$ is only approximately satisfied (Fig. 1b), our numerical examples still exhibits decaying asymptotic variance in the neighbourhood of the critical point. In this case, however, the asymptotic variance diverges for growing values of the perturbation μ. If the perturbations are opposed ($\mu = -\nu$), it is possible for certain observables that the unperturbed dynamics represents a global minimum. Such a case is observed in Fig. 1c. In Fig. 1d, e the observable $f(q) = l \cdot q$ is considered. If the damping is sufficiently strong ($\gamma > \sqrt{2}$), the unperturbed dynamics is at a local maximum of the asymptotic variance (Fig. 1d). Furthermore, the asymptotic variance approaches zero as $\mu \to \infty$ (for a theoretical explanation see again Sect. 4.3). The graph in Fig. 1e shows that the assumption of γ not being too small cannot be dropped from Theorem 3. Even in this case though the example shows decay of the asymptotic variance for large values of μ.

4.2 Exponential Decay Rate

Let us denote by λ^* the *optimal exponential decay rate* in (21), i.e.

$$\lambda^* = \sup\{\lambda > 0 \,|\, \text{There exists } C \geq 1 \text{ such that (21) holds}\}. \tag{42}$$

Note that λ^* is well-defined and positive by Theorem 2. We also define the *spectral bound* of the generator $\mathcal{L}$ by

$$s(\mathcal{L}) = \inf(\mathrm{Re}\,\sigma(-\mathcal{L}) \setminus \{0\}). \tag{43}$$

In [38] it is proven that the Ornstein–Uhlenbeck semigroup $(P_t)_{t\geq 0}$ considered in this section is differentiable (see Proposition 2.1). In this case (see Corollary 3.12 of [17]), it is known that the exponential decay rate and the spectral bound coincide, i.e. $\lambda^* = s(\mathcal{L})$, whereas in general only $\lambda^* \leq s(\mathcal{L})$ holds. In this section we will therefore analyse the spectral properties of the generator $\mathcal{L}$ in the Gaussian case. In particular, this leads to some intuition of why choosing equal perturbations ($\mu = \nu$) is crucial for the performance of the sampler.

In [38] (see also [43]), it was proven that the spectrum of $\mathcal{L}$ as in (36) in $L^2(\widehat{\pi})$ is given by

$$\sigma(\mathcal{L}) = \left\{ -\sum_{j=1}^{r} n_j \lambda_j \,:\, n_j \in \mathbb{N}, \lambda_j \in \sigma(B) \right\}. \tag{44}$$

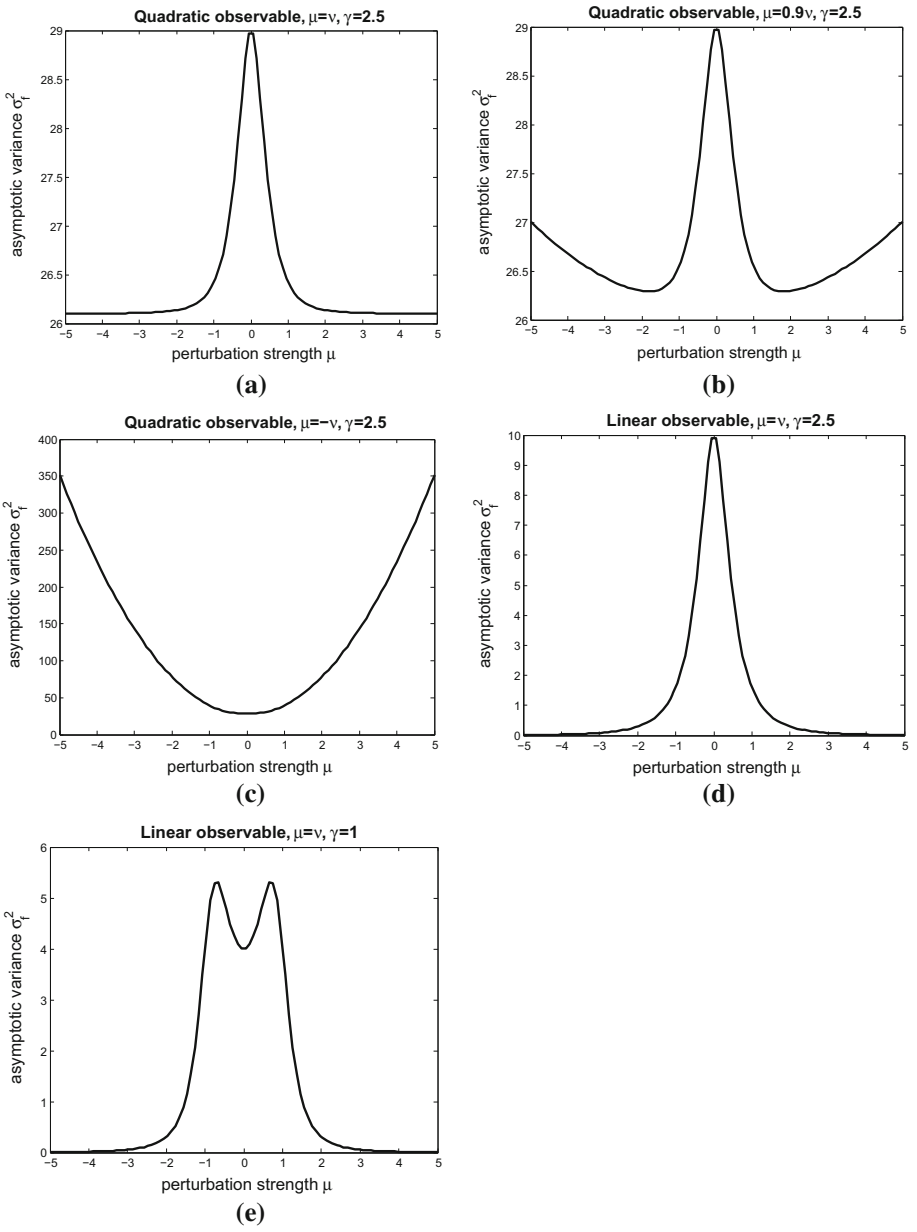

Fig. 1 Asymptotic variance for linear and quadratic observables, depending on relative perturbation and friction strengths. **a** Equal perturbations: $\mu = \nu$. **b** Approximately equal perturbations: $\mu = 0.9\nu$. **c** Opposing perturbations: $\mu = -\nu$. **d** Equal perturbations: $\mu = \nu$ (sufficiently large friction γ). **e** Equal perturbations: $\mu = \nu$ (small friction γ)

Note that $\sigma(\mathcal{L})$ only depends on the drift matrix B. In the case where $\mu = \nu$, the spectrum of B can be computed explicitly.

Lemma 4 *Assume $\mu = \nu$. Then the spectrum of B is given by*[2]

$$\sigma(B) = \left\{\mu\lambda + \sqrt{\left(\frac{\gamma}{2}\right)^2 - 1} + \frac{\gamma}{2} | \lambda \in \sigma(J)\right\} \cup \left\{\mu\lambda - \sqrt{\left(\frac{\gamma}{2}\right)^2 - 1} + \frac{\gamma}{2} | \lambda \in \sigma(J)\right\}. \quad (45)$$

Proof We will compute $\sigma\left(B - \frac{\gamma}{2}I\right)$ and then use the identity $\sigma(B) = \left\{\lambda + \frac{\gamma}{2} | \lambda \in \sigma\left(B - \frac{\gamma}{2}I\right)\right\}$. We have

$$\begin{aligned}\det\left(B - \frac{\gamma}{2}I - \lambda I\right) &= \det\left(\left(\mu J - \frac{\gamma}{2}I - \lambda I\right)\left(\mu J + \frac{\gamma}{2}I - \lambda I\right) + I\right) \\ &= \det\left(\mu J - \lambda I + \sqrt{\left(\frac{\gamma}{2}\right)^2 - 1} I\right) \cdot \det\left(\mu J - \lambda I - \sqrt{\left(\frac{\gamma}{2}\right)^2 - 1} I\right),\end{aligned}$$

where I is understood to denote the identity matrix of appropriate dimension. The above quantity is zero if and only if

$$\lambda - \sqrt{\left(\frac{\gamma}{2}\right)^2 - 1} \in \sigma(\mu J) \quad \text{or} \quad \lambda + \sqrt{\left(\frac{\gamma}{2}\right)^2 - 1} \in \sigma(\mu J).$$

Together with (4.2), the claim follows. □

Using formula (45), in Fig. 2a we show a sketch of the spectrum $\sigma(-\mathcal{L})$ for the case of equal perturbations ($\mu = \nu$) with the convenient choices $n = 1$ and $\gamma = 2$. Of course, the eigenvalue at 0 is associated to the invariant measure since $\mathcal{L}^\dagger \widehat{\pi} = 0$. The arrows indicate the movement of the eigenvalues as the perturbation μ increases in accordance with Lemma 4. Clearly, the spectral bound of $\mathcal{L}$ is not affected by the perturbation. Note that the eigenvalues on the real axis stay invariant under the perturbation. The subspace of $L_0^2(\widehat{\pi})$ associated to those will turn out to be crucial for the characterisation of the limiting asymptotic variance as $\mu \to \infty$ (see Remark 10).

To illustrate the suboptimal properties of the perturbed dynamics when the perturbations are not equal, we plot the spectrum of the drift matrix $\sigma(B)$ in the case when the dynamics is only perturbed by the term $\nu J_2 p \mathrm{d}t$ (i.e. $\mu = 0$) for $n = 2$, $\gamma = 2$ and

$$J_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad (46)$$

(see Fig. 2b). Note that the full spectrum $\sigma(-\mathcal{L})$ can be inferred from (44). For $\nu = 0$ we have that the spectrum $\sigma(B)$ only consists of the (degenerate) eigenvalue 1. For increasing ν, the figure shows that the degenerate eigenvalue splits up into four eigenvalues, two of which get closer to the imaginary axis as ν increases, leading to a smaller spectral bound and therefore to a decrease in the speed of convergence to equilibrium. Figure 2a, b give an intuitive explanation of why the fine-tuning of the perturbation strengths is crucial.

We close this subsection by providing autocorrelation plots (see Fig. 3) for the linear observable considered in Fig. 1d (with a friction coefficient of $\gamma = 2.5$). It is well-known that the asymptotic variance is given by the integrated autocorrelation function (see e.g. Proposition IV 1.3 in [3]),

$$\sigma_f^2 = \int_0^\infty \mathbb{E}_\pi \left[(f(q_0) - \pi(f))(f(q_t) - \pi(f))\right] \mathrm{d}t. \quad (47)$$

[2] Notice that $\sqrt{\left(\frac{\gamma}{2}\right)^2 - 1}$ is understood to be a complex number for $\gamma < 2$.

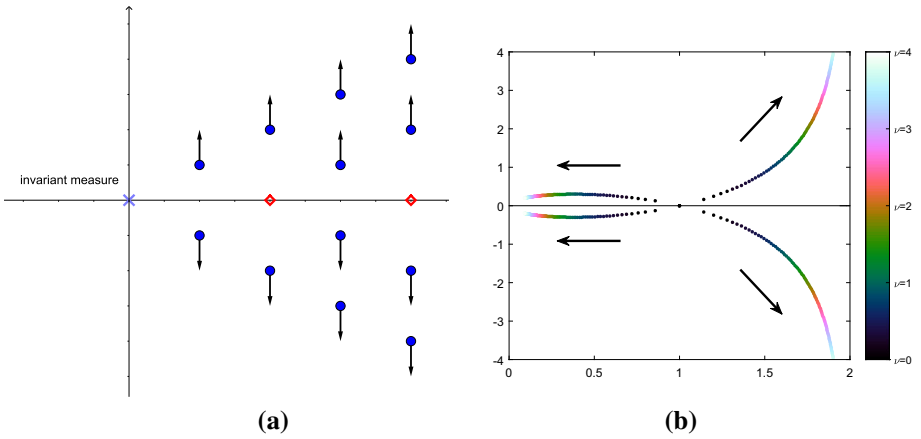

Fig. 2 Effects of the perturbation on the spectra of $-\mathcal{L}$ and B. **a** $\sigma(-\mathcal{L})$ in the case $\mu = \nu$. The arrows indicate the movement of the spectrum as the perturbation strength μ increases. **b** $\sigma(B)$ in the case $J_1 = 0$, i.e. the dynamics is only perturbed by $-\nu J_2 p \mathrm{d}t$. The arrows indicate the movement of the eigenvalues as ν increases

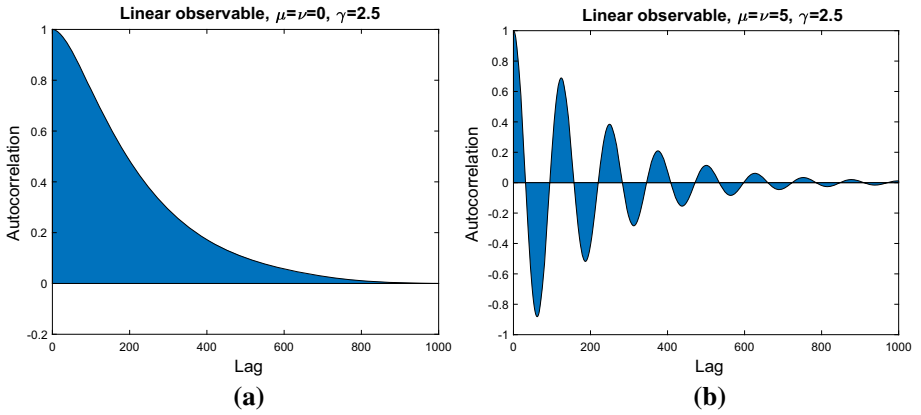

Fig. 3 Autocorrelation plots for the perturbed and unperturbed dynamics. **a** Unperturbed Langevin dynamics. **b** Perturbed Langevin dynamics

Comparing Fig. 3a, b yields additional insight into the mechanics of the variance reduction: the increase of the imaginary part of the eigenvalues of $\mathcal{L}$ (as indicated in Fig. 2a) leads to oscillations in the autocorrelation function and therefore to cancellations in (47). A similar effect has already been observed in [50] for the nonreversible overdamped Langevin dynamics (15).

4.3 Unit Covariance: Large Perturbations

In the previous subsection we observed that for the particular perturbation $J_1 = J_2$ and $\mu = \nu$ [see equation (34)] the perturbed Langevin dynamics demonstrated an improvement in performance for μ in a neighbourhood of 0, when the observable is linear or quadratic. Recall that this dynamics is ergodic with respect to a standard Gaussian measure $\hat{\pi}$ on $\mathbb{R}^{2d}$ with marginal π with respect to the q-variable. As before, we shall consider only observables that

do not depend on p. Moreover, we assume without loss of generality that $\pi(f) = 0$. For such observables we will write $f \in L^2_0(\pi)$ and consider the canonical embedding $L^2_0(\pi) \subset L^2(\hat{\pi})$. We emphasize that $L^2_0(\pi)$ consists of functions that only depend on q, whereas functions in $L^2(\hat{\pi})$ may depend on both q and p.

In this subsection will analyse the asymptotic variance for large values of μ. The infinitesimal generator of (34) can be written as

$$\mathcal{L} = \underbrace{p \cdot \nabla_q - q \cdot \nabla_p + \gamma(-p \cdot \nabla_p + \Delta_p)}_{\mathcal{L}_0} + \mu \underbrace{(-Jq \cdot \nabla_q - Jp \cdot \nabla_p)}_{\mathcal{A}} =: \mathcal{L}_0 + \mu\mathcal{A}, \quad (48)$$

where we have introduced the notation $\mathcal{L}_{pert} = \mu\mathcal{A}$. In the sequel, the adjoint of an operator $\mathcal{B}$ in $L^2(\widehat{\pi})$ will be denoted by $\mathcal{B}^*$. In the rest of this section we will make repeated use of the Hermite polynomials

$$g_\alpha(x) = (-1)^{|\alpha|} e^{\frac{|x|^2}{2}} \nabla^\alpha e^{-\frac{|x|^2}{2}}, \quad \alpha \in \mathbb{N}^{2d}, \quad (49)$$

invoking the notation $x = (q, p) \in \mathbb{R}^{2d}$. For $m \in \mathbb{N}_0$ define the Hilbert spaces

$$H_m = \text{span}\{g_\alpha : |\alpha| = m\}, \quad \langle f, g\rangle_m := \langle f, g\rangle_{L^2(\widehat{\pi})}, \quad f, g \in H_m.$$

The following result (Theorem 4) holds for operators of the form (36) providing an orthogonal decomposition of $L^2(\widehat{\pi})$ into invariant subspaces. The drift and diffusion matrices B and Q are assumed to be such that $\mathcal{L}$ is the generator of an ergodic stochastic process (see [2, Definition 2.1] for precise conditions).

Theorem 4 [2, Sect. 5]. *The following holds:*

(a) *The space $L^2(\widehat{\pi})$ has a decomposition into mutually orthogonal subspaces:*

$$L^2(\widehat{\pi}) = \bigoplus_{m\in\mathbb{N}_0} H_m.$$

(b) *For all $m \in \mathbb{N}_0$, H_m is invariant under $\mathcal{L}$ as well as under the semigroup $(e^{t\mathcal{L}})_{t\geq 0}$.*
(c) *The spectrum of $\mathcal{L}$ has the following decomposition:*

$$\sigma(\mathcal{L}) = \bigcup_{m\in\mathbb{N}_0} \sigma(\mathcal{L}|_{H_m}), \quad \sigma(\mathcal{L}|_{H_m}) = \left\{ \sum_{j=1}^{2d} \alpha_j \lambda_j : |\alpha| = m,\ \lambda_j \in \sigma(B) \right\}.$$

Remark 6 Note that by the ergodicity of the dynamics, $\ker \mathcal{L}$ consists of constant functions and so $\ker \mathcal{L} = H_0$. Therefore, $L^2_0(\widehat{\pi})$ has the decomposition

$$L^2_0(\widehat{\pi}) = L^2(\widehat{\pi})/\ker \mathcal{L} = \bigoplus_{m\geq 1} H_m.$$

Our first main result of this section is an expression for the asymptotic variance in terms of the unperturbed operator $\mathcal{L}_0$ and the perturbation $\mathcal{A}$:

Proposition 2 *Let $f \in L^2_0(\pi)$ (so in particular $f = f(q)$). Then the associated asymptotic variance is given by*

$$\sigma_f^2 = \langle f, -\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1} f\rangle_{L^2(\widehat{\pi})}. \quad (50)$$

Remark 7 The proof of the preceding Proposition will show that $\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A}$ is invertible on $L^2_0(\widehat{\pi})$ and that $(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1} f \in \mathcal{D}(\mathcal{L}_0)$ for all $f \in L^2_0(\widehat{\pi})$.

To prove Proposition 2 we will make use of the *generator with reversed perturbation*

$$\mathcal{L}_- = \mathcal{L}_0 - \mu\mathcal{A}$$

and the *momentum flip operator*

$$P : L_0^2(\widehat{\pi}) \to L_0^2(\widehat{\pi}), \quad \phi(q, p) \mapsto \phi(q, -p). \tag{51}$$

Clearly, $P^2 = I$ and $P^* = P$. Further properties of $\mathcal{L}_0$, $\mathcal{A}$ and the auxiliary operators $\mathcal{L}_-$ and P are gathered in the following lemma:

Lemma 5 *For all $\phi, \psi \in C^\infty(\mathbb{R}^{2d}) \cap L^2(\widehat{\pi})$ the following holds:*

(a) *The generator $\mathcal{L}_0$ is symmetric in $L^2(\widehat{\pi})$ with respect to P:*

$$\langle \phi, P\mathcal{L}_0 P\psi\rangle_{L^2(\widehat{\pi})} = \langle \mathcal{L}_0\phi, \psi\rangle_{L^2(\widehat{\pi})}.$$

(b) *The perturbation $\mathcal{A}$ is skewadjoint in $L^2(\widehat{\pi})$:*

$$\mathcal{A}^* = -\mathcal{A}.$$

(c) *The operators $\mathcal{L}_0$ and $\mathcal{A}$ commute:*

$$[\mathcal{L}_0, \mathcal{A}]\phi = 0.$$

(d) *The perturbation $\mathcal{A}$ satisfies*

$$P\mathcal{A}P\phi = \mathcal{A}\phi.$$

(e) *$\mathcal{L}$ and $\mathcal{L}_-$ commute,*

$$[\mathcal{L}, \mathcal{L}_-]\phi = 0,$$

and the following relation holds:

$$\langle \phi, P\mathcal{L}P\psi\rangle_{L^2(\widehat{\pi})} = \langle \mathcal{L}_-\phi, \psi\rangle_{L^2(\widehat{\pi})}. \tag{52}$$

(f) *The operators $\mathcal{L}$, $\mathcal{L}_0$, $\mathcal{L}_-$, $\mathcal{A}$ and P leave the Hermite spaces H_m invariant.*

Remark 8 The claim (c) in the above lemma is crucial for our approach, which itself rests heavily on the fact that the $q-$ and $p-$perturbations match ($J_1 = J_2$).

Proof of Lemma 5 The statement (a) is well-known and its proof can be found in [35, Sect. 2.2.3.1] for instance. The claim (b) follows by noting that the flow vector field $b(q, p) = (-Jq, -Jp)$ associated to $\mathcal{A}$ is divergence-free with respect to $\widehat{\pi}$, i.e. $\nabla\cdot(\widehat{\pi}b) = 0$. Therefore, $\mathcal{A}$ is the generator of a strongly continuous unitary semigroup on $L^2(\widehat{\pi})$ and hence skewadjoint by Stone's Theorem. The claims (c), (d) and (e) follow by direct computations which can be found in [41]. To prove (f) first notice that $\mathcal{L}$, $\mathcal{L}_0$ and $\mathcal{L}_-$ are of the form (36) and therefore leave the spaces H_m invariant by Theorem 4. It follows immediately that also $\mathcal{A}$ leaves those spaces invariant. The fact that P leaves the spaces H_m invariant follows directly by inspection of (49) and (51). □

Now we proceed with the proof of Proposition 2:

Proof of Proposition 2 Since the potential V is quadratic, Assumption 1 clearly holds and thus Lemma 2 ensures that $\mathcal{L}$ and $\mathcal{L}_-$ are invertible on $L^2_0(\widehat{\pi})$ with

$$(-\mathcal{L})^{-1} = \int_0^\infty e^{t\mathcal{L}} \mathrm{d}t, \tag{53}$$

and analogously for $\mathcal{L}_-^{-1}$. In particular, the asymptotic variance can be written as $\sigma_f^2 = \langle f, (-\mathcal{L})^{-1} f\rangle_{L^2(\widehat{\pi})}$. Due to the respresentation (53) and Theorem 4, the inverses of $\mathcal{L}$ and $\mathcal{L}_-$ leave the Hermite spaces H_m invariant. We will prove the claim from Proposition 2 under the assumption that $Pf = f$ which includes the case $f = f(q)$. For the following calculations we will assume $f \in H_m$ for fixed $m \geq 1$. Combining statement (f) with (a) and (e) of Lemma 5 (and noting that $H_m \subset C^\infty(\mathbb{R}^{2d}) \cap L^2(\widehat{\pi})$) we see that

$$P\mathcal{L}P = \mathcal{L}_-^* \quad \text{and} \quad P\mathcal{L}_0 P = \mathcal{L}_0^* \tag{54}$$

when restricted to H_m. Therefore, the following calculations are justified:

$$\begin{aligned}
\langle f, (-\mathcal{L})^{-1} f\rangle_{L^2(\widehat{\pi})} &= \frac{1}{2}\langle f, (-\mathcal{L})^{-1} f\rangle_{L^2(\widehat{\pi})} + \langle f, (-\mathcal{L}^*)^{-1} f\rangle_{L^2(\widehat{\pi})} \\
&= \frac{1}{2}\langle f, (-\mathcal{L})^{-1} f\rangle_{L^2(\widehat{\pi})} + \langle Pf, (-\mathcal{L}^*)^{-1} Pf\rangle_{L^2(\widehat{\pi})} \\
&= \frac{1}{2}\langle f, (-\mathcal{L})^{-1} f\rangle_{L^2(\widehat{\pi})} + \langle f, (-\mathcal{L}_-)^{-1} f\rangle_{L^2(\widehat{\pi})} \\
&= \frac{1}{2}\langle f, \left((-\mathcal{L})^{-1} + (-\mathcal{L}_-)^{-1}\right) f\rangle_{L^2(\widehat{\pi})},
\end{aligned}$$

where in the second line we have used the assumption $Pf = f$ and in the third line the properties $P^2 = I$, $P^* = P$ and Eq. (54). Since $\mathcal{L}$ and $\mathcal{L}_-$ commute on H_m according to Lemma 5(e),(f) we can write

$$(-\mathcal{L})^{-1} + (-\mathcal{L}_-)^{-1} = \mathcal{L}_-(-\mathcal{L}\mathcal{L}_-)^{-1} + \mathcal{L}(-\mathcal{L}\mathcal{L}_-)^{-1} = -2\mathcal{L}_0(\mathcal{L}\mathcal{L}_-)^{-1}$$

for the restrictions on H_m, using $\mathcal{L}+\mathcal{L}_- = 2\mathcal{L}_0$. We also have $\mathcal{L}\mathcal{L}_- = (\mathcal{L}_0+\mu\mathcal{A})(\mathcal{L}_0-\mu\mathcal{A}) = \mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A}$, since $\mathcal{L}_0$ and $\mathcal{A}$ commute. We thus arrive at the formula

$$\sigma_f^2 = \langle f, -\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1} f\rangle_{L^2(\widehat{\pi})}, \quad f \in H_m. \tag{55}$$

Now since $(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1} f = (\mathcal{L}\mathcal{L}_-)^{-1} f \in \mathcal{D}(\mathcal{L}_0)$ for all $f \in L^2(\widehat{\pi})$, it follows that the operator $-\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1}$ is bounded. We can therefore extend formula (55) to the whole of $L^2(\widehat{\pi})$ by continuity, using the fact that $L^2_0(\widehat{\pi}) = \bigoplus_{m\geq 1} H_m$. □

Applying Proposition 2 we can analyse the behaviour of σ_f^2 in the limit of large perturbation strength $\mu \to \infty$. To this end, we introduce the orthogonal decomposition

$$L^2_0(\pi) = \ker(Jq\cdot\nabla_q) \oplus \ker(Jq\cdot\nabla_q)^\perp, \tag{56}$$

where $Jq\cdot\nabla_q$ is understood as an unbounded operator acting on $L^2_0(\pi)$, obtained as the smallest closed extension of $Jq\cdot\nabla_q$ acting on $C_c^\infty(\mathbb{R}^d)$. In particular, $\ker(Jq\cdot\nabla_q)$ is a closed linear subspace of $L^2_0(\pi)$. Let Π denote the $L^2_0(\pi)$-orthogonal projection onto $\ker(Jq\cdot\nabla_q)$. We will write $\sigma_f^2(\mu)$ to stress the dependence of the asymptotic variance on the perturbation strength. The following result shows that for large perturbations, the limiting asymptotic variance is always smaller than the asymptotic variance in the unperturbed case. Furthermore, the limit is given as the asymptotic variance of the projected observable Πf for the unperturbed dynamics.

Theorem 5 *Let $f \in L_0^2(\pi)$ (so in particular $f = f(q)$). Then* $\lim_{\mu\to\infty} \sigma_f^2(\mu) = \sigma_{\Pi f}^2(0) \leq \sigma_f^2(0)$.

Remark 9 Note that the fact that the limit exists and is finite is nontrivial. In particular, as Fig. 1b, c demonstrate, it is often the case that $\lim_{\mu\to\infty} \sigma_f^2(\mu) = \infty$ if the condition $\mu = \nu$ is not satisfied.

Remark 10 The decomposition (56) can be interpreted in terms of the spectrum $\sigma(\mathcal{L})$ as follows: First observe that for functions f that only depend on q, $f \in \ker(Jq \cdot \nabla_q)$ is equivalent to $f \in \ker \mathcal{A}$. Let us denote by $\bar{\sigma} = \bigcap_{\mu\in\mathbb{R}} \sigma(\mathcal{L}_0 + \mu\mathcal{A})$ the part of $\sigma(\mathcal{L}_0)$ that is not affected by the perturbation and by

$$\bar{E} = \overline{\operatorname{span}\{f \in L^2(\hat{\pi}) : \text{ there exists } \lambda \in \bar{\sigma} \text{ such that } \mathcal{L}_0 f = \lambda f\}} \subset L^2(\hat{\pi})$$

the corresponding subspace. Then it is straightforward to see that $\ker(\mathcal{A}) = \bar{E}$.[3] In Fig. 2a, $\bar{\sigma}$ has been highlighted by diamonds.

Proof of Theorem 5 Note that $\mathcal{L}_0$ and $\mathcal{A}^*\mathcal{A}$ leave the Hermite spaces H_m invariant and their restrictions to those spaces commute (see Lemma 5, (b), (c) and (f)). Furthermore, as the Hermite spaces H_m are finite-dimensional, those operators have discrete spectrum. As $\mathcal{A}^*\mathcal{A}$ is nonnegative self-adjoint, there exists an orthogonal decomposition $L_0^2(\pi) = \bigoplus_i W_i$ into eigenspaces of the operator $-\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1}$, the decomposition $\bigoplus W_i$ being finer then $\bigoplus H_m$ in the sense that every W_i is a subspace of some H_m. Moreover, $-\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\mathcal{A}^*\mathcal{A})^{-1}|_{W_i} = -\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\lambda_i)^{-1}|_{W_i}$, where $\lambda_i \geq 0$ is the eigenvalue of $\mathcal{A}^*\mathcal{A}$ associated to the subspace W_i. Consequently, formula (50) can be written as

$$\sigma_f^2 = \sum_i \langle f_i, -\mathcal{L}_0(\mathcal{L}_0^2 + \mu^2\lambda_i)^{-1} f_i\rangle_{L^2(\hat{\pi})}, \tag{57}$$

where $f = \sum_i f_i$ and $f_i \in W_i$. Let us assume now without loss of generality that $W_0 = \ker \mathcal{A}^*\mathcal{A}$, so in particular $\lambda_0 = 0$. Then clearly

$$\lim_{\mu\to\infty} \sigma_f^2 = 2\langle f_0, -\mathcal{L}_0(\mathcal{L}_0^2)^{-1} f_0\rangle_{L^2(\hat{\pi})} = 2\langle f_0, (-\mathcal{L}_0)^{-1} f_0\rangle_{L^2(\hat{\pi})} = \sigma_{f_0}^2(0).$$

Now notice that $W_0 = \ker \mathcal{A}^*\mathcal{A} = \ker \mathcal{A}$, showing the equality in the claim. It remains to show that $\sigma_{\Pi f}^2(0) \leq \sigma_f^2(0)$. To see this, we write

$$\begin{aligned}\sigma_f^2(0) &= 2\left\langle f, (-\mathcal{L}_0)^{-1} f\right\rangle_{L^2(\hat{\pi})} = 2\langle \Pi f + (1-\Pi)f, (-\mathcal{L}_0)^{-1}\left(\Pi f + (1-\Pi)f\right)\rangle_{L^2(\hat{\pi})} \\ &= \sigma_{\Pi f}^2(0) + \sigma_{(1-\Pi)f}^2(0) + R,\end{aligned}$$

where

$$R = 2\langle \Pi f, (-\mathcal{L}_0)^{-1}(1-\Pi)f\rangle_{L^2(\hat{\pi})} + 2\langle (1-\Pi)f, (-\mathcal{L}_0)^{-1}\Pi f\rangle_{L^2(\hat{\pi})}.$$

Note that since we only consider observables that do not depend on p, $\Pi f \in \ker(Jq \cdot \nabla_q)$ and $(1-\Pi)f \in \bigoplus_{i\geq 1} W_i$. Since $\mathcal{L}_0$ commutes with $\mathcal{A}$, it follows that $(-\mathcal{L}_0)^{-1}$ leaves both W_0 and $\bigoplus_{i\geq 1} W_i$ invariant. Therefore, as the latter spaces are orthogonal to each other, it follows that $R = 0$, from which the result follows. □

[3] Indeed, the fact that $f \in \ker \mathcal{A}$ is equivalent to $f \in \bar{E}$ is easy to check if f is an eigenvector of $\mathcal{L}_0$ (recall that f is then an eigenvector of $\mathcal{L}_0 + \mu\mathcal{A}$ as well, using Lemma 5(c) The claim then follows by extending linearly.

From Theorem 5 it follows that in the limit as $\mu \to \infty$, the asymptotic variance $\sigma_f^2(\mu)$ is not decreased by the perturbation if $f \in \ker(Jq \cdot \nabla_q)$. In fact, this result also holds true non-asymptotically, i.e. observables in $\ker(Jq \cdot \nabla_q)$ are not affected at all by the perturbation:

Lemma 6 *Let $f \in \ker(Jq \cdot \nabla_q)$. Then $\sigma_f^2(\mu) = \sigma_f^2(0)$ for all $\mu \in \mathbb{R}$.*

Proof From $f \in \ker(Jq \cdot \nabla_q)$ it follows immediately that $f \in \ker \mathcal{A}^*\mathcal{A}$. Then the claim follows from the expression (57). □

Example 1 Recall the case of observables of the form $f(q) = q \cdot Kq + l \cdot q + C$ with $K \in \mathbb{R}^{d\times d}_{sym}$, $l \in \mathbb{R}^d$ and $C \in \mathbb{R}$ from Sect. 4.1. If $[J, K] = 0$ and $l \in \ker J$, then $f \in \ker(Jq \cdot \nabla_q)$ as

$$Jq \cdot \nabla_q(q \cdot Kq + l \cdot q + C) = 2Jq \cdot Kq + Jq \cdot l = q \cdot (KJ - JK)q - q \cdot Jl = 0.$$

From the preceding lemma it follows that $\sigma_f^2(\mu) = \sigma_f^2(0)$ for all $\mu \in \mathbb{R}$, showing that the assumption in Theorem 3 does not exclude nontrivial cases.

The following result shows that the dynamics (34) is particularly effective for antisymmetric observables (at least in the limit of large perturbations):

Proposition 3 *Let $f \in L_0^2(\pi)$ satisfy $f(-q) = -f(q)$ and assume that $\ker J = \{0\}$. Furthermore, assume that the eigenvalues of J are rationally independent, i.e.*

$$\sigma(J) = \{\pm i\lambda_1, \pm i\lambda_2, \ldots, \pm i\lambda_d\} \tag{58}$$

with $\lambda_i \in \mathbb{R}_{>0}$ and $\sum_i k_i\lambda_i \neq 0$ for all $(k_1, \ldots, k_d) \in \mathbb{Z}^d \setminus (0, \ldots, 0)$. Then $\lim_{\mu\to\infty} \sigma_f^2(\mu) = 0$.

Proof of Proposition 3 The claim would immediately follow from $f \in \ker(Jq \cdot \nabla)^\perp$ according to Theorem 5, but that does not seem to be so easy to prove directly. Instead, we again make use of the Hermite polynomials.

Recall from the proof of Proposition 2 that $\mathcal{L}$ is invertible on $L_0^2(\widehat{\pi})$ and its inverse leaves the Hermite spaces H_m invariant. Consequently, the asymptotic variance of an observable $f \in L_0^2(\widehat{\pi})$ can be written as

$$\sigma_f^2 = \langle f, (-\mathcal{L})^{-1} f\rangle_{L^2(\widehat{\pi})} = \sum_{m=1}^{\infty} \langle \Pi_m f, (-\mathcal{L}|_{H_m})^{-1} \Pi_m f\rangle_{L^2(\widehat{\pi})}, \tag{59}$$

where $\Pi_m : L_0^2(\widehat{\pi}) \to H_m$ denotes the orthogonal projection onto H_m. From (49) it is clear that g_a is symmetric for $|\alpha|$ even and antisymmetric for $|\alpha|$ odd. Therefore, from f being antisymmetric it follows that $f \in \bigoplus_{m\geq 1, m \text{ odd}} H_m$. In view of (45), ((c)) and (58) the spectrum of $\mathcal{L}_{|H_m}$ can be written as

$$\begin{aligned}\sigma(\mathcal{L}|_{H_m}) &= \left\{ \mu \sum_{j=1}^{2d} \alpha_j\beta_j + C_{\alpha,\gamma} : |\alpha| = m,\ \beta_j \in \sigma(J) \right\} \\ &= \left\{ i\mu \sum_{j=1}^{d} (\alpha_j - \alpha_{j+d})\lambda_j + C_{\alpha,\gamma} : |\alpha| = m \right\}\end{aligned} \tag{60}$$

with appropriate real constants $C_{\alpha,\gamma} \in \mathbb{R}$ that depend on α and γ, but not on μ. For $|\alpha| = \sum_{j=1}^{2d} \alpha_j = m$ odd, we have that

$$\sum_{j=1}^{d} (\alpha_j - \alpha_{j+d})\lambda_j \neq 0. \tag{61}$$

Indeed, assume to the contrary that the above expression is zero. Then it would follow that $\alpha_j = \alpha_{j+d}$ for all $j = 1, \ldots, d$ by rational independence of $\lambda_1, \ldots, \lambda_d$ and $|m|$ would have to be even. From (60) and (61) it is clear that

$$\sup\left\{r > 0 : B(0,r) \cap \sigma(\mathcal{L}|_{H_m}) = \emptyset\right\} \xrightarrow{\mu\to\infty} \infty,$$

where $B(0,r)$ denotes the ball of radius r centered at the origin in $\mathbb{C}$. Consequently, the spectral radius of $(-\mathcal{L}|_{H_m})^{-1}$ and hence $(-\mathcal{L}|_{H_m})^{-1}$ itself converges to zero as $\mu \to \infty$. The result then follows from (59). □

Remark 11 The idea of the preceding proof can be explained using Fig. 2a and Remark 10. The eigenvalues in the fixed spectrum $\bar{E}$ (on the real axis, highlighted by diamonds) correspond to Hermite polynomials of even order. The independence condition on the eigenvalues of J prevents cancellations that would lead to fixed eigenvalues associated to Hermite polynomials of odd order. Therefore, antisymmetric observables are orthogonal to $\bar{E} = \ker \mathcal{A}$.

The following corollary gives a version of the converse of Proposition 3 and provides further intuition into the mechanics of the variance reduction achieved by the perturbation.

Corollary 1 *Let $f \in L_0^2(\pi)$ and assume that $lim_{\mu\to\infty}\sigma_f^2(\mu) = 0$. Then*

$$\int_{B(0,r)} f \mathrm{dq} = 0$$

for all $r \in (0,\infty)$, where $B(0,r)$ denotes the ball centered at 0 with radius r.

Proof According to Theorem 5, $\lim_{\mu\to\infty} \sigma_f^2(\mu) = 0$ implies $\sigma_{\Pi f}^2(0) = 0$. We can write

$$\sigma_{\Pi f}^2(0) = \langle \Pi f, (-\mathcal{L}_0)^{-1}\Pi f\rangle_{L^2(\widehat{\pi})} = \frac{1}{2}\langle \Pi f, \left((-\mathcal{L}_0)^{-1} + (-\mathcal{L}_0^*)^{-1}\right)\Pi f\rangle_{L^2(\widehat{\pi})} \tag{62}$$

and recall from the proof of Proposition 2 that $(-\mathcal{L}_0)^{-1}$ and $(-\mathcal{L}_0^*)^{-1}$ leave the Hermite spaces H_m invariant. Therefore $\ker\left((-\mathcal{L}_0)^{-1} + (-\mathcal{L}_0^*)^{-1}\right) = 0$ in $L_0^2(\widehat{\pi})$, and in particular $\sigma_{\Pi f}^2(0) = 0$ implies $\Pi f = 0$, which in turn shows that $f \in \ker(Jq\cdot\nabla)^\perp$. From $(Jq\cdot\nabla)^* = -Jq\cdot\nabla$, it follows that

$$\ker(Jq\cdot\nabla)^\perp = \overline{\mathrm{im}(Jq\cdot\nabla)^*} = \overline{\mathrm{im}(Jq\cdot\nabla)}. \tag{63}$$

Hence, there exists a sequence $(\phi_n)_n \in C_c^\infty(\mathbb{R}^d)$ such that $Jq\cdot\nabla\phi_n \to f$ in $L^2(\pi)$. Taking a subsequence if necessary, we can assume that the convergence is pointwise π-almost everywhere and that the sequence is pointwise bounded by a function in $L^1(\pi)$. Since J is antisymmetric, we have that $Jq\cdot\nabla\phi_n = \nabla\cdot(\phi_n Jq)$. Now Gauss's theorem yields

$$\int_{B(0,r)} f \mathrm{dq} = \int_{B(0,r)} \nabla\cdot(\phi Jq)\mathrm{dq} = \int_{\partial B(0,r)} \phi Jq\cdot \mathrm{dn},$$

where n denotes the outward normal to the sphere $\partial B(0,r)$. This quantity is zero due to the orthogonality of Jq and n, and so the result follows from Lebesgue's dominated convergence theorem. □

4.4 Optimal Choices of J for Quadratic Observables

Assume $f \in L^2_0(\pi)$ is given by $f(q) = q \cdot Kq + l \cdot q - \operatorname{Tr} K$, with $K \in \mathbb{R}^{d\times d}_{sym}$ and $l \in \mathbb{R}^d$ (note that the constant term is chosen such that $\pi(f) = 0$). Our objective is to choose J in such a way that $\lim_{\mu\to\infty} \sigma^2_f(\mu)$ becomes as small as possible. To stress the dependence on the choice of J, we introduce the notation $\sigma^2_f(\mu, J)$. Also, we denote the orthogonal projection onto $(\ker J)^\perp$ by $\Pi^\perp_{\ker J}$.

Lemma 7 (Zero variance limit for linear observables). *Assume $K = 0$ and $\Pi^\perp_{\ker J} l = 0$. Then*

$$\lim_{\mu\to\infty} \sigma^2_f(\mu, J) = 0.$$

Proof According to Theorem 5, we have to show that $\Pi f = 0$, where Π is the $L^2(\pi)$-orthogonal projection onto $\ker(Jq \cdot \nabla)$. Let us thus use (63) and prove that $f \in \overline{\operatorname{im}(Jq \cdot \nabla)}$. Indeed, since $\Pi^\perp_{\ker J} l = 0$, by Fredholm's alternative there exists $u \in \mathbb{R}^d$ such that $Ju = l$. Now define $\phi \in L^2_0(\pi)$ by $\phi(q) = -u \cdot q$, leading to $f = Jq \cdot \nabla\phi$, so the result follows. □

Lemma 8 (Zero variance limit for purely quadratic observables.) *Let $l = 0$ and consider the decomposition $K = K_0 + K_1$ into the traceless part $K_0 = K - \frac{\operatorname{Tr} K}{d} \cdot I$ and the trace-part $K_1 = \frac{\operatorname{Tr} K}{d} \cdot I$. For the corresponding decomposition of the observable*

$$f(q) = f_0(q) + f_1(q) = q \cdot K_0 q + q \cdot K_1 q - \operatorname{Tr} K$$

the following holds:

(a) *There exists an antisymmetric matrix J such that $\lim_{\mu\to\infty} \sigma^2_{f_0}(\mu, J) = 0$, and there is an algorithmic way (see Algorithm 1) to compute an appropriate J in terms of K.*
(b) *The trace-part is not effected by the perturbation, i.e. $\sigma^2_{f_1}(\mu, J) = \sigma^2_{f_1}(0)$ for all $\mu \in \mathbb{R}$.*

Proof To prove the first claim, according to Theorem 5 it is sufficient to show that $f_0 \in \ker(Jq \cdot \nabla)^\perp = \overline{\operatorname{im}(Jq \cdot \nabla)}$. Let us consider the function $\phi(q) = q \cdot Aq$, with $A \in \mathbb{R}^{d\times d}_{sym}$. It holds that $Jq \cdot \nabla\phi = 2q \cdot (J^T Aq) = q \cdot [A, J]q$. The task of finding an antisymmetric matrix J such that $\lim_{\mu\to\infty} \sigma^2_{f_0}(\mu, J) = 0$ can therefore be accomplished by constructing an antisymmetric matrix J such that there exists a symmetric matrix A with the property $K_0 = [A, J]$. Given any traceless matrix K_0 there exists an orthogonal matrix $U \in O(\mathbb{R}^d)$ such that UK_0U^T has zero entries on the diagonal, and that U can be obtained in an algorithmic manner (see for example [29] or [22, Chap. 2, Sect. 2, Problem 3]) Assume thus that such a matrix $U \in O(\mathbb{R}^d)$ has been found and choose real numbers $a_1, \ldots, a_d \in \mathbb{R}$ such that $a_i \neq a_j$ if $i \neq j$. We now set $\bar{A} = \operatorname{diag}(a_1, \ldots, a_n)$, and

$$\bar{J}_{ij} = \begin{cases} \frac{(UK_0U^T)_{ij}}{a_i - a_j} & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases} \tag{64}$$

Observe that since UK_0U^T is symmetric, $\bar{J}$ is antisymmetric. A short calculation shows that $[\bar{A}, \bar{J}] = UK_0U^T$. We can thus define $A = U^T \bar{A} U$ and $J = U^T \bar{J} U$ to obtain $[A, J] = K_0$. Therefore, the J constructed in this way indeed satisfies (4.4). For the second claim, note that $f_1 \in \ker(Jq \cdot \nabla)$, since $Jq \cdot \nabla \left(q \cdot \frac{\operatorname{Tr} K}{d} q\right) = 2\frac{\operatorname{Tr} K}{d} q \cdot Jq = 0$ due to the antisymmetry of J. The result then follows from Lemma 6. □

We would like to stress that the perturbation J constructed in the previous lemma is far from unique due to the freedom of choice of U and $a_1, \ldots, a_d \in \mathbb{R}$ in its proof. However, it is asymptotically optimal:

Corollary 2 *In the setting of Lemma 8 the following holds:*

$$\min_{J^T=-J}\left(\lim_{\mu\to\infty}\sigma_f^2(\mu,J)\right)=\sigma_{f_1}^2(0).$$

Proof The claim follows immediately since $f_1 \in \ker(Jq\cdot\nabla)$ for arbitrary antisymmetric J as shown in (4.4), and therefore the contribution of the trace part f_1 to the asymptotic variance cannot be reduced by any choice of J according to Lemma 6. □

As the proof of Lemma 8 is constructive, we obtain the following algorithm for determining optimal perturbations for quadratic observables:

Algorithm 1 *Given $K \in \mathbb{R}^{d\times d}_{sym}$, determine an optimal antisymmetric perturbation J as follows:*

1. *Set $K_0 = K - \frac{\mathrm{Tr}\,K}{d}\cdot I$.*
2. *Find $U \in O(\mathbb{R}^d)$ such that UK_0U^T has zero entries on the diagonal.*
3. *Choose $a_i \in \mathbb{R}$, $i = 1,\dots d$ such that $a_i \neq a_j$ for $i \neq j$ and set*

$$\bar{J}_{ij} = \frac{(UK_0U^T)_{ij}}{a_i - a_j}$$

   for $i \neq j$ and $\bar{J}_{ii} = 0$ otherwise.
4. *Set $J = U^T\bar{J}U$.*

Remark 12 In [14], the authors consider the task of finding optimal perturbations J for the nonreversible overdamped Langevin dynamics given in (15). In the Gaussian case this optimization problem turns out be equivalent to the one considered in this section. Indeed, equation (39) of [14] can be rephrased as $f \in \ker(Jq\cdot\nabla)^\perp$. Therefore, Algorithm 1 and its generalization Algorithm 2 (described in Sect. 4.5) can be used without modifications to find optimal perturbations of overdamped Langevin dynamics.

4.5 Gaussians with Arbitrary Covariance and Preconditioning

In this section we extend the results of the preceding sections to the case when the target measure π is given by a Gaussian with arbitrary covariance, i.e. $V(q) = \frac{1}{2}q\cdot Sq$ with $S \in \mathbb{R}^{d\times d}_{sym}$ symmetric and positive definite. The dynamics (8) then takes the form

$$\mathrm{d}q_t = M^{-1}p_t\mathrm{d}t - \mu J_1 S q_t\mathrm{d}t,\quad \mathrm{d}p_t = -Sq_t\mathrm{d}t - \nu J_2 M^{-1}p_t\mathrm{d}t - \Gamma M^{-1}p_t\mathrm{d}t + \sqrt{2\Gamma}\mathrm{d}W_t. \tag{65}$$

The key observation is now that the choices $M = S$ and $\Gamma = \gamma S$ together with the transformation $\widetilde{q} = S^{1/2}q$ and $\widetilde{p} = S^{-1/2}p$ lead to the dynamics

$$\begin{aligned}\mathrm{d}\widetilde{q}_t &= \widetilde{p}_t\mathrm{d}t - \mu S^{1/2}J_1S^{1/2}\widetilde{q}_t\mathrm{d}t,\\ \mathrm{d}\widetilde{p}_t &= -\widetilde{q}_t\mathrm{d}t - \mu S^{-1/2}J_2S^{-1/2}\widetilde{p}_t\mathrm{d}t - \gamma\widetilde{p}_t\mathrm{d}t + \sqrt{2\gamma}\mathrm{d}W_t,\end{aligned} \tag{66}$$

which is of the form (34) if J_1 and J_2 obey the condition $SJ_1S = J_2$. Clearly the dynamics (66) is ergodic with respect to a Gaussian measure with unit covariance, in the following denoted by $\widetilde{\pi}$. The connection between the asymptotic variances associated to (65) and (66) is as follows:

For an observable $f \in L_0^2(\pi)$ we can write

$$\sqrt{T}\left(\frac{1}{T}\int_0^T f(q_s)\mathrm{d}s - \pi(f)\right) = \sqrt{T}\left(\frac{1}{T}\int_0^T \widetilde{f}(\widetilde{q}_s)\mathrm{d}s - \widetilde{\pi}(\widetilde{f})\right),$$

where $\widetilde{f}(q) = f(S^{-1/2}q)$. Therefore, the asymptotic variances satisfy $\sigma_f^2 = \widetilde{\sigma}_{\widetilde{f}}^2$, where $\widetilde{\sigma}_{\widetilde{f}}^2$ denotes the asymptotic variance of the process $(\widetilde{q}_t)_{t\geq 0}$. Because of this, the results from the previous sections generalise to (65), subject to the condition that the choices $M = S$, $\Gamma = \gamma S$ and $SJ_1S = J_2$ are made. We formulate our results in this general setting as corollaries:

Corollary 3 *Consider the dynamics*

$$\begin{aligned}\mathrm{d}q_t &= M^{-1}p_t\mathrm{d}t - \mu J_1\nabla V(q_t)\mathrm{d}t,\\ \mathrm{d}p_t &= -\nabla V(q_t)\mathrm{d}t - \mu J_2M^{-1}p_t\mathrm{d}t - \Gamma M^{-1}p_t\mathrm{d}t + \sqrt{2\Gamma}\mathrm{d}W_t,\end{aligned} \tag{67}$$

with $V(q) = \frac{1}{2}q\cdot Sq$. Assume that $M = S$, $\Gamma = \gamma S$ with $\gamma > \sqrt{2}$ and $SJ_1S = J_2$. Let $f \in L^2(\pi)$ be an observable of the form

$$f(q) = q\cdot Kq + l\cdot q + C \tag{68}$$

with $K \in \mathbb{R}^{d\times d}_{sym}$, $l \in \mathbb{R}^d$ and $C \in \mathbb{R}$. If at least one of the conditions $KJ_1S \neq SJ_1K$ and $l \notin \ker J$ is satisfied, then the asymptotic variance is at a local maximum for the unperturbed sampler, i.e.

$$\partial_\mu\sigma_f^2\Big|_{\mu=0} = 0 \qquad \text{and} \qquad \partial_\mu^2\sigma_f^2\Big|_{\mu=0} < 0.$$

Proof Note that

$$\widetilde{f}(q) = f(S^{-1/2}q) = q\cdot S^{-1/2}KS^{-1/2}q + S^{-1/2}l\cdot q + C = q\cdot\widetilde{K}q + \widetilde{l}\cdot q + C$$

is again of the form (68) (where in the last equality, $\widetilde{K} = S^{-1/2}KS^{-1/2}$ and $\widetilde{l} = S^{-1/2}l$ have been defined). From (66), (4.5) and Theorem 3 the claim follows if at least one of the conditions $[\widetilde{K}, S^{1/2}J_1S^{1/2}] \neq 0$ and $\widetilde{l} \notin \ker(S^{1/2}J_1S^{1/2})$ is satisfied. The first of those can easily seen to be equivalent to $S^{-1/2}(KJS - SJK)S^{-1/2} \neq 0$, which is equivalent to $KJ_1S \neq SJ_1K$ since S is nondegenerate. The second condition is equivalent to $S^{1/2}J_1l \neq 0$, which is equivalent to $J_1l \neq 0$, again by nondegeneracy of S. □

Corollary 4 *Assume the setting from the previous corollary and denote by Π the orthogonal projection onto* $\ker(J_1Sq\cdot\nabla)$. *For $f \in L^2(\pi)$ it holds that* $\lim_{\mu\to\infty}\sigma_f^2(\mu) = \sigma_{\Pi f}^2(0) \leq \sigma_f^2(0)$.

Proof Theorem 5 implies $\lim_{\mu\to\infty}\widetilde{\sigma}_{\widetilde{f}}^2(\mu) = \widetilde{\sigma}_{\widetilde{\Pi}\widetilde{f}}^2(0) \leq \widetilde{\sigma}_{\widetilde{f}}^2(0)$ for the transformed system (66). Here $\widetilde{f}(q) = f(S^{-1/2}q)$ is the transformed observable and $\widetilde{\Pi}$ denotes $L^2(\pi)$-orthogonal projection onto $\ker(S^{1/2}J_1S^{1/2}q\cdot\nabla)$. According to (4.5), it is sufficient to show that $(\Pi f)\circ S^{-1/2} = \widetilde{\Pi}\widetilde{f}$. This however follows directly from the fact that the linear transformation $\phi \mapsto \phi\circ S^{1/2}$ maps $\ker(S^{1/2}J_1S^{1/2}q\cdot\nabla)$ bijectively onto $\ker(J_1Sq\cdot\nabla)$. □

Let us also reformulate Algorithm 1 for the case of a Gaussian with arbitrary covariance.

Algorithm 2 *Given $K, S \in \mathbb{R}^{d\times d}_{sym}$ with $f(q) = q\cdot Kq$ and $V(q) = \frac{1}{2}q\cdot Sq$ (assuming S is nondegenerate), determine optimal perturbations J_1 and J_2 as follows:*

1. *Set $\widetilde{K} = S^{-1/2}KS^{-1/2}$ and $\widetilde{K}_0 = \widetilde{K} - \frac{\mathrm{Tr}\,\widetilde{K}}{d}\cdot I$.*

2. *Find* $U \in O(\mathbb{R}^d)$ *such that* $U\widetilde{K}_0 U^T$ *has zero entries on the diagonal.*
3. *Choose* $a_i \in \mathbb{R}$, $i = 1, \ldots, d$ *such that* $a_i \neq a_j$ *for* $i \neq j$ *and set*

$$\bar{J}_{ij} = \frac{(U\widetilde{K}_0 U^T)_{ij}}{a_i - a_j}.$$

4. *Set* $\widetilde{J} = U^T \bar{J} U$.
5. *Put* $J_1 = S^{-1/2}\widetilde{J}S^{-1/2}$ *and* $J_2 = S^{1/2} J S^{1/2}$.

Finally, we obtain the following optimality result from Lemma 7 and Corollary 2.

Corollary 5 *Let* $f(q) = q \cdot Kq + l \cdot q - \operatorname{Tr} K$ *and assume that* $\Pi^{\perp}_{\ker J} l = 0$. *Then*

$$\min_{J_1^T = -J_1,\, J_2 = SJ_1S} \left(\lim_{\mu \to \infty} \sigma_f^2(\mu, J_1, J_2) \right) = \sigma_{f_1}^2(0),$$

where $f_1(q) = q \cdot K_1 q$, $K_1 = \frac{\operatorname{Tr}(S^{-1}K)}{d} S$. *Optimal choices for* J_1 *and* J_2 *can be obtained using Algorithm 2.*

Remark 13 Since in Sect. 4.1 we analysed the case where J_1 and J_2 are proportional, we are not able to drop the restriction $J_2 = SJ_1S$ from the above optimality result. Analysis of completely arbitrary perturbations will be the subject of future work.

Remark 14 The choices $M = S$ and $\Gamma = \gamma S$ have been introduced to make the perturbations considered in this article lead to samplers that perform well in terms of reducing the asymptotic variance. However, adjusting the mass and friction matrices according to the target covariance in this way (i.e. $M = S$ and $\Gamma = \gamma S$) is a popular way of preconditioning the dynamics, see for instance [18] and, in particular mass-tensor molecular dynamics [6]. Here we will present an argument why such a preconditioning is indeed beneficial in terms of the convergence rate of the dynamics. Let us first assume that S is diagonal, i.e. $S = \operatorname{diag}(s^{(1)}, \ldots, s^{(d)})$ and that $M = \operatorname{diag}(m^{(d)}, \ldots, m^{(d)})$ and $\Gamma = \operatorname{diag}(\gamma^{(d)}, \ldots, \gamma^{(d)})$ are chosen diagonally as well. Then (65) decouples into one-dimensional SDEs of the following form:

$$\mathrm{d}q_t^{(i)} = \frac{1}{m^{(i)}} p_t^{(i)} \mathrm{d}t, \quad \mathrm{d}p_t^{(i)} = -s^{(i)} q_t^{(i)} \mathrm{d}t - \frac{\gamma^{(i)}}{m^{(i)}} p_t^{(i)} \mathrm{d}t + \sqrt{2\gamma^{(i)}} \mathrm{d}W_t, \quad i = 1, \ldots, d. \tag{69}$$

Let us write those Ornstein–Uhlenbeck processes as

$$\mathrm{d}X_t^{(i)} = -B^{(i)} X_t^{(i)} \mathrm{d}t + \sqrt{2Q^{(i)}} \mathrm{d}W_t^{(i)}, \quad B^{(i)} = \begin{pmatrix} 0 & -\frac{1}{m^{(i)}} \\ s^{(i)} & \frac{\gamma^{(i)}}{m^{(i)}} \end{pmatrix}, \quad Q^{(i)} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma^{(i)} \end{pmatrix}. \tag{70}$$

As in Sect. 4.2, the rate of the exponential decay of (70) is equal to $\min \operatorname{Re} \sigma(B^{(i)})$. A short calculation shows that the eigenvalues of $B^{(i)}$ are given by

$$\lambda_{1,2}^{(i)} = \frac{\gamma^{(i)}}{2m^{(i)}} \pm \sqrt{\left(\frac{\gamma^{(i)}}{2m^{(i)}}\right)^2 - \frac{s^{(i)}}{m^{(i)}}}.$$

Therefore, the rate of exponential decay is maximal when

$$\left(\frac{\gamma^{(i)}}{2m^{(i)}}\right)^2 - \frac{s^{(i)}}{m^{(i)}} = 0, \tag{71}$$

in which case it is given by

$$(\lambda^{(i)})^* = \sqrt{\frac{s^{(i)}}{m^{(i)}}}.$$

Naturally, it is reasonable to choose $m^{(i)}$ in such a way that the exponential rate $(\lambda^{(i)})^*$ is the same for all i, leading to the restriction $M = cS$ with $c > 0$. Choosing c small will result in fast convergence to equilibrium, but also make the dynamics (69) quite stiff, requiring a very small timestep Δt in a discretisation scheme. The choice of c will therefore need to strike a balance between those two competing effects. The constraint (71) then implies $\Gamma = 2cS$. By a coordinate transformation, the preceding argument also applies if S, M and Γ are diagonal in the same basis, and of course M and Γ can always be chosen that way. Numerical experiments show that it is possible to increase the rate of convergence to equilibrium even further by choosing M and Γ nondiagonally with respect to S (although only by a small margin). A clearer understanding of this is a topic of further investigation.

5 Numerical Experiments: Diffusion Bridge Sampling

5.1 Numerical Scheme

In this section we introduce a splitting scheme for simulating the perturbed underdamped Langevin dynamics given by Eq. (8). In the unpertubed case, i.e. when $J_1 = J_2 = 0$, the $BAOAB$ scheme (see [33] and references therein) has proven to be efficient for computing long time ergodic averages with respect to q-dependent observables. Motivated by this, we introduce the following perturbed scheme, introducing additional Runge-Kutta integration steps:

$$p_{n+1/2} = p_n - \frac{1}{2}\Delta t \nabla V(q_n), \tag{72a}$$

$$q_{n+1/2} = q_n + \frac{1}{2}\Delta t \cdot M^{-1} p_{n+1/2}, \tag{72b}$$

$$q'_{n+1/2} = RK_4\left(\frac{1}{2}\Delta t, q_{n+1/2}\right), \tag{72c}$$

$$\hat{p} = \exp(-\Delta t(\Gamma M^{-1} + \nu J_2 M^{-1}))p_{n+1/2} + \sqrt{I - e^{-2\Gamma \Delta t}}\mathcal{N}(0,1), \tag{72d}$$

$$q''_{n+1/2} = RK_4\left(\frac{1}{2}\Delta t, q'_{n+1/2}\right), \tag{72e}$$

$$q_{n+1} = q''_{n+1/2} + \frac{1}{2}\Delta t \cdot M^{-1}\hat{p}, \tag{72f}$$

$$p_{n+1} = \hat{p} - \frac{1}{2}\Delta t \cdot \nabla V(q_{n+1}), \tag{72g}$$

where $RK_4(\Delta t, q_0)$ refers to fourth order Runge-Kutta integration of the ODE

$$\dot{q} = -J_1 \nabla V(q), \quad q(0) = q_0 \tag{73}$$

up until time Δt. We remark that the J_2-perturbation is linear and can therefore be included in the O-part without much computational overhead. We emphasize the fact that many different splitting schemes could be investigated: although the BAOAB-scheme works well for unperturbed Langevin dynamics, it is not clear whether this remains true for the perturbed

dynamics. Moreover, the perturbations introduced by J_1 and J_2 can be added in various places. Other discretisation schemes for the ODE (73) could be useful as well, for instance one could use a symplectic integrator, using the Hamiltonian structure of (73). However, since V as the Hamiltonian for (73) is not separable in general, such a symplectic integrator would have to be implicit. Note that (72c) and (72e) could be merged since (72e) commutes with (72d). In this paper, we content ourselves with the above scheme for our numerical experiments. Investigation of optimal numerical schemes for perturbed Langevin dynamics is an interesting problem for further research.

Remark 15 The aformentioned schemes lead to an error in the approximation for $\pi(f)$, since the invariant measure π is not preserved exactly by the numerical scheme. In practice, the $BAOAB$-scheme can therefore be accompanied by an accept-reject Metropolis step as in [40], leading to an unbiased estimate of $\pi(f)$, albeit with an inflated variance. In this case, after every rejection the momentum variable has to be flipped ($p \mapsto -p$) in order to keep the correct invariant measure. We note here that our perturbed scheme can be 'Metropolized' in a similar way by 'flipping' the matrices J_1 and J_2 after every rejection ($J_1 \mapsto -J_1$ and $J_2 \mapsto -J_2$) and using an appropriate (volume-preserving and time-reversible) integrator for the dynamics given by (73). Implementations of this idea are the subject of ongoing work. See [47] for a similar approach to nonreversible overdamped Langevin dynamics.

5.2 Diffusion Bridge Sampling

To numerically test our analytical results, we will apply the dynamics (8) to sample a measure on path space associated to a diffusion bridge. Specifically, consider the SDE

$$\mathrm{d}X_s = -\nabla U(X_s)\mathrm{d}s + \sqrt{2\beta^{-1}}\mathrm{d}W_s,$$

with $X_s \in \mathbb{R}^n$, $\beta > 0$ and the potential $U : \mathbb{R}^n \to \mathbb{R}$ obeying adequate growth and smoothness conditions (see [24], Sect. 5 for precise statements). The law of the solution to this SDE conditioned on the events $X(0) = x_-$ and $X(s_+) = x_+$ is a probability measure π on $L^2([0, s_+], \mathbb{R}^n)$ which poses a challenging and important sampling problem, especially if U is multimodal. This setting has been used as a test case for sampling probability measures in high dimensions (see for example [9] and [45]). For a more detailed introduction (including applications) see [11] and for a rigorous theoretical treatment the papers [11,24–26].

In the case $U \equiv 0$, it can be shown that the law of the conditioned process is given by a Gaussian measure π_0 with mean zero and precision operator $\mathcal{S} = -\frac{\beta}{2}\Delta$ on the Sobolev space $H^1([0, s_+], \mathbb{R}^d)$ equipped with appropriate boundary conditions. The general case can then be understood as a perturbation thereof: The measure π is absolutely continuous with respect to π_0 with Radon-Nikodym derivative

$$\frac{\mathrm{d}\pi}{\mathrm{d}\pi_0} \propto \exp\big(-\Psi\big), \tag{74}$$

where

$$\Psi(x) = \frac{\beta}{2}\int_0^{s_+} G(x(s), \beta)\mathrm{d}s \quad \text{and} \quad G(x, \beta) = \frac{1}{2}|\nabla U(x)|^2 - \frac{1}{\beta}\Delta U(x).$$

We will make the choice $x_- = x_+ = 0$, which is possible without loss of generality as explained in [10, Remark 3.1], leading to Dirichlet boundary conditions on $[0, s_+]$ for the

precision operator $\mathcal{S}$. Furthermore, we choose $s_+ = 1$ and discretise the ensuing s-interval $[0, 1]$ according to

$$[0, 1] = [0, s_1) \cup [s_1, s_2) \cup \ldots \cup [s_{n-1}, s_n) \cup [s_n, 1]$$

in an equidistant way with stespize $s_{j+1} - s_j \equiv \delta = \frac{1}{d+1}$. Functions on this grid are determined by the values $x(s_1) = x_1, \ldots, x(s_n) = x_n$, recalling that $x(0) = x(1) = 0$ by the Dirichlet boundary conditions. We discretise the functional Ψ as

$$\tilde{\Psi}(x_1, \ldots, x_n) = \frac{\beta}{2}\delta \sum_{i=1}^{d} G(x_i, \beta) = \frac{\beta}{2}\delta \sum_{i=1}^{d} \Big(\frac{1}{2}(U'(x_i)^2 - \frac{1}{\beta}U''(x_i)\Big),$$

such that its gradient is given by

$$(\nabla\tilde{\Psi})_i = \frac{\beta}{2}\delta\big(2U'(x_i)U''(x_i) - \frac{1}{\beta}U'''(x_i)\big), \quad i = 1, \ldots, d.$$

We denote by A_δ the discretised Dirichlet Laplacian on $[0, 1]$ with stepsize δ. Following (74), the discretised target measure $\widehat{\pi}$ has the form

$$\widehat{\pi} = \frac{1}{Z}e^{-V}\mathrm{d}x \quad \text{with} \quad V(x) = \tilde{\Psi}(x) - \frac{\beta\delta}{4}x \cdot A_\delta x, \quad x \in \mathbb{R}^d.$$

In the following we will consider the case $n = 1$ with potential $U : \mathbb{R} \to \mathbb{R}$ given by $U(x) = \frac{1}{2}(x^2 - 1)^2$ and set $\beta = 1$. To test our algorithm we adjust the parameters M, Γ, J_1 and J_2 according to the recommended choice in the Gaussian case, (27), where we take $S = \frac{\beta}{2}\delta \cdot A_\delta$ as the precision operator of the Gaussian target. We will consider the linear observable $f_1(x) = l \cdot x$ with $l = (1, \ldots, 1)$ and the quadratic observable $f_2(x) = |x|^2$. In a first experiment we adjust the perturbations J_1 and J_2 to the observable f_2 according to Algorithm 2. The dynamics (8) is integrated using the splitting scheme introduced in Sect. 5.1 with a stepsize of $\Delta t = 10^{-4}$ over the time interval $[0, T]$ with $T = 10^2$. Furthermore, we choose initial conditions $q_0 = (1, \ldots, 1)$, $p_0 = (0, \ldots, 0)$ and introduce a burn-in time $T_0 = 1$, i.e. we take the estimator to be $\hat{\pi}(f) \approx \frac{1}{T-T_0}\int_{T_0}^{T} f(q_t)\mathrm{d}t$. We compute the variance of the above estimator from $N = 500$ realisations and compare the results for different choices of the friction coefficient γ and of the perturbation strength μ.

The numerical experiments show that the perturbed dynamics generally outperform the unperturbed dynamics independently of the choice of μ and γ, both for linear and quadratic observables. One notable exception is the behaviour of the linear observable for small friction $\gamma = 10^{-3}$ (see Fig. 4a), where the asymptotic variance initially increases for small perturbation strengths μ. This does not contradict our analytical results, since the small perturbation results Theorem 3 and Corollary 3 are only valid if $\gamma > \sqrt{2}$. We remark here that the condition $\gamma > \sqrt{2}$, while necessary for the theoretical results from Sect. 4.1, is not a very advisable choice in practice (at least in this experiment), since Figs. 4b and 5b clearly indicate that the optimal friction is around $\gamma \approx 10^{-1}$. Interestingly, the problem of choosing a suitable value for the friction coefficient coefficient γ becomes mitigated by the introduction of the perturbation: While the performance of the unperturbed sampler depends quite sensitively on γ, the asymptotic variance of the perturbed dynamics is a lot more stable with respect to variations of γ. A somewhat surprising phenomenon is that the standard deviation σ associated to the linear observable decays in the range $\gamma \in [10, 100]$ for the unperturbed sampler (see Fig. 4b). We confirmed this behaviour by further numerical experiments and remark that as the target measure $\hat{\pi}$ is fairly complicated, convexity of the function $\gamma \mapsto \sigma$ should not be expected.

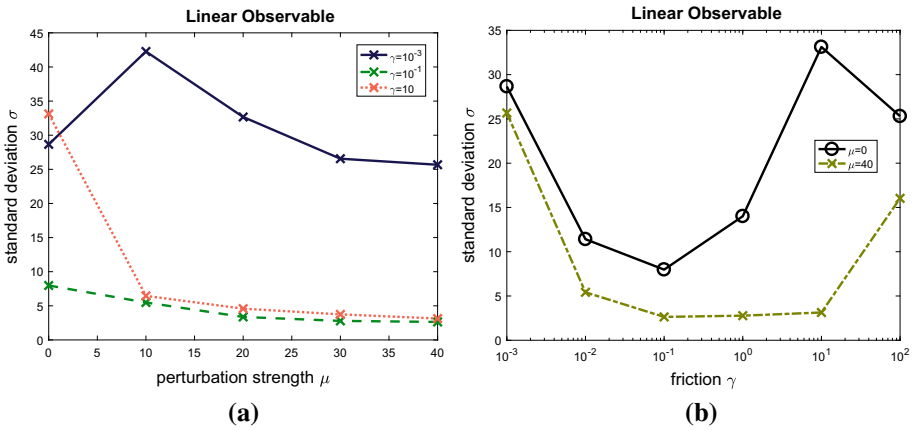

Fig. 4 Standard deviation of $\hat{\pi}(f)$ for a linear observable as a function of friction γ and perturbation strength μ

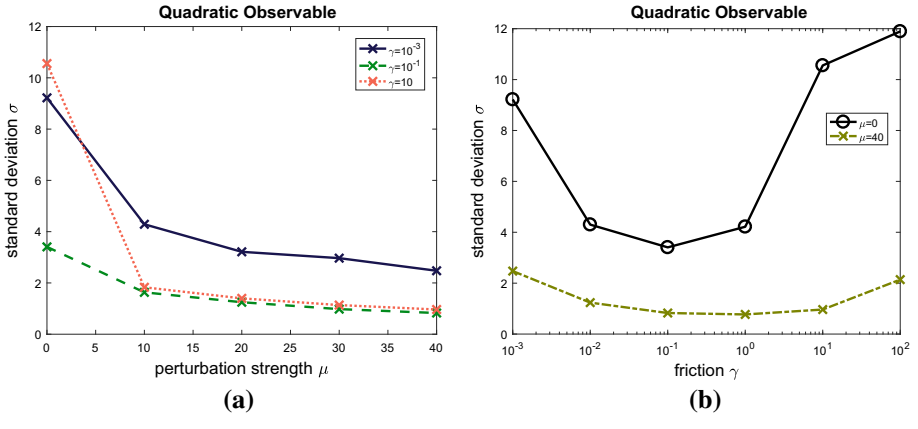

Fig. 5 Standard deviation of $\hat{\pi}(f)$ for a quadratic observable as a function of friction γ and perturbation strength μ

In the regime of growing values of μ, the experiments confirm the results from Sect. 4.3, i.e. the asymptotic variance approaches a limit that is smaller than the asymptotic variance of the unperturbed dynamics.

As a final remark we report our finding that the performance of the sampler for the linear observable is qualitatively independent of the choice of J_1 [as long as J_2 is adjusted according to (27)]. This result is in alignment with Propostion 3 which predicts good properties of the sampler for antisymmetric observables. In contrast to this, a judicious choice of J_1 is critical for quadratic observables. In particular, applying Algorithm 2 significantly improves the performance of the perturbed sampler in comparison to choosing J_1 arbitrarily.

6 Outlook and Future Work

A new family of Langevin samplers was introduced in this paper. These new SDE samplers consist of perturbations of the underdamped Langevin dynamics (that is known to be ergodic

with respect to the canonical measure), where auxiliary drift terms in the equations for both the position and the momentum are added, in a way that the perturbed family of dynamics is ergodic with respect to the same (canonical) distribution. These new Langevin samplers were studied in detail for Gaussian target distributions where it was shown, using tools from spectral theory for differential operators, that an appropriate choice of the perturbations in the equations for the position and momentum can improve the performance of the Langvin sampler, at least in terms of reducing the asymptotic variance. The performance of the perturbed Langevin sampler to non-Gaussian target densities was tested numerically on the problem of diffusion bridge sampling.

The work presented in this paper can be improved and extended in several directions. First, a rigorous analysis of the new family of Langevin samplers for non-Gaussian target densities is needed. The analytical tools developed in [14] can be used as a starting point. Furthermore, the study of the actual computational cost and its minimization by an appropriate choice of the numerical scheme and of the perturbations in position and momentum would be of interest to practitioners. In addition, the analysis of our proposed samplers can be facilitated by using tools from symplectic and differential geometry. Finally, combining the new Langevin samplers with existing variance reduction techniques such as zero variance MCMC, preconditioning/Riemannian manifold MCMC can lead to sampling schemes that can be of interest to practitioners, in particular in molecular dynamics simulations. All these topics are currently under investigation.

Acknowledgements AD was supported by the EPSRC under Grant No. EP/J009636/1. NN is supported by EPSRC through a Roth Departmental Scholarship. GP is partially supported by the EPSRC under Grants No. EP/J009636/1, EP/L024926/1, EP/L020564/1 and EP/L025159/1. Part of the work reported in this paper was done while NN and GP were visiting the Institut Henri Poincaré during the Trimester Program "Stochastic Dynamics Out of Equilibrium". The hospitality of the Institute and of the organizers of the program is greatly acknowledged. We thank the referees for their useful comments and suggestions that have lead to various improvements in the presentation of this paper.

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. Achleitner, F., Arnold, A., Stürzer, D.: Large-time behavior in non-symmetric Fokker-Planck equations. Riv. Math. Univ. Parma (N.S.) **6**(1), 1–68 (2015)
2. Arnold, A., Erb, J.: Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift. arXiv:1409.5425v2 (2014)
3. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis. Stochastic Modelling and Applied Probability, vol. 57. Springer, New York (2007)
4. Alrachid, H., Mones, L., Ortner, C.: Some remarks on preconditioning molecular dynamics. arXiv preprint arXiv:1612.05435 (2016)
5. Bass, R.F.: Diffusions and Elliptic Operators. Springer, Berlin (1998)
6. Bennett, C.H.: Mass tensor molecular dynamics. J. Comput. Phys. **19**(3), 267–279 (1975)
7. Bakry, D., Gentil, I., Ledoux, M.: Analysis and Geometry of Markov Diffusion Operators, vol. 348. Springer, Berlin (2013)
8. Bhattacharya, R.N.: On the functional central limit theorem and the law of the iterated logarithm for Markov processes. Z. Wahrsch. Verw. Gebiete **60**(2), 185–201 (1982)
9. Beskos, A., Pinski, F.J., Sanz-Serna, J.M., Stuart, A.M.: Hybrid Monte Carlo on Hilbert spaces. Stoch. Process. Appl. **121**(10), 2201–2230 (2011)

10. Beskos, A., Roberts, G., Stuart, A., Voss, J.: MCMC methods for diffusion bridges. Stoch. Dyn. **8**(3), 319–350 (2008)
11. Beskos, A., Stuart, A.: MCMC methods for sampling function space. In: ICIAM 07—6th International Congress on Industrial and Applied Mathematics, pp. 337–364. European Mathematical Society, Zürich (2009)
12. Ceriotti, M., Bussi, G., Parrinello, M.: Langevin equation with colored noise for constant-temperature molecular dynamics simulations. Phys. Rev. Lett. **102**(2), 020601 (2009)
13. Cattiaux, P., Chafaı, D., Guillin, A.: Central limit theorems for additive functionals of ergodic markov diffusions processes. ALEA **9**(2), 337–382 (2012)
14. Duncan, A.B., Lelievre, T., Pavliotis, G.A.: Variance reduction using nonreversible Langevin samplers. J. Stat. Phys. **163**(3), 457–491 (2016)
15. Dolbeault, J., Mouhot, C., Schmeiser, C.: Hypocoercivity for linear kinetic equations conserving mass. Trans. Am. Math. Soc. **367**(6), 3807–3828 (2015)
16. Eyink, G.L., Lebowitz, J.L., Spohn, H.: Hydrodynamics and fluctuations outside of local equilibrium: driven diffusive systems. J. Stat. Phys. **83**(3–4), 385–472 (1996)
17. Engel, K.-J., Nagel, R.: One-parameter semigroups for linear evolution equations, volume 194 of Graduate Texts in Mathematics. Springer, New York (2000). With contributions by Brendle, S., Campiti, M., Hahn, T., Metafune, G., Nickel, G., Pallara, D., Perazzoli, C., Rhandi, A., Romanelli, S., Schnaubelt, R
18. Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. J. R. Stat. Soc. Ser. B Stat. Methodol. **73**(2), 123–214 (2011). With discussion and a reply by the authors
19. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis. Texts in Statistical Science Series, 3rd edn. CRC Press, Boca Raton, FL (2014)
20. Hwang, C.-R., Hwang-Ma, S., Sheu, S.J.: Accelerating Gaussian diffusions. Ann. Appl. Probab. **3**(3), 897–913 (1993)
21. Hwang, C.-R., Hwang-Ma, S.-Y., Sheu, S.-J.: Accelerating diffusions. Ann. Appl. Probab. **15**(2), 1433–1444 (2005)
22. Horn, Roger A., Johnson, Charles R.: Matrix Analysis, 2nd edn. Cambridge University Press, Cambridge (2013)
23. Hwang, C.-R., Normand, R., Wu, S.-J.: Variance reduction for diffusions. Stoch. Process. Appl. **125**(9), 3522–3540 (2015)
24. Hairer, M., Stuart, A.M., Voss, J.: Analysis of SPDEs arising in path sampling. II. The nonlinear case. Ann. Appl. Probab. **17**(5–6), 1657–1706 (2007)
25. Hairer, M., Stuart, A., Voss, J.: Sampling conditioned diffusions. In: Trends in Stochastic Analysis. London London Mathematical Society Lecture Note Series, vol. 353, pp. 159–185. Cambridge University Press, Cambridge (2009)
26. Hairer, M., Stuart, A.M., Voss, J., Wiberg, P.: Analysis of SPDEs arising in path sampling. I. The Gaussian case. Commun. Math. Sci. **3**(4), 587–603 (2005)
27. Iacobucci, A., Olla, S., Stoltz, G.: Convergence rates for nonequilibrium Langevin dynamics. arXiv:1702.03685 (2017)
28. Joulin, A., Ollivier, Y.: Curvature, concentration and error estimates for Markov chain Monte Carlo. Ann. Probab. **38**(6), 2418–2442 (2010)
29. Kazakia, J.Y.: Orthogonal transformation of a trace free symmetric matrix into one with zero diagonal elements. Int. J. Eng. Sci. **26**(8), 903–906 (1988)
30. Kliemann, W.: Recurrence and invariant measures for degenerate diffusions. Ann Probab **15**, 690–707 (1987)
31. Komorowski, T., Landim, C., Olla, S.: Fluctuations in Markov processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 345. Springer, Heidelberg. Time symmetry and martingale approximation (2012)
32. Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer, Berlin (2008)
33. Leimkuhler, B., Matthews,C.: Molecular Dynamics. Interdisciplinary Applied Mathematics, vol. 39. Springer, Berlin (2015). With deterministic and stochastic numerical methods
34. 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)
35. Lelièvre, T., Rousset, M., Stoltz, G.: Free Energy Computations. Imperial College Press, London (2010). A Mathematical Perspective
36. Lelièvre, T., Stoltz, G.: Partial differential equations and stochastic methods in molecular dynamics. Acta Numer. **25**, 681–880 (2016)
37. 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)

38. Metafune, G., Pallara, D., Priola, E.: Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures. J. Funct. Anal. **196**(1), 40–60 (2002)
39. Markowich, P.A., Villani, C.: On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. Mat. Contemp. **19**, 1–29 (2000)
40. Matthews, C., Weare, J., Leimkuhler, B.: Ensemble preconditioning for Markov Chain Monte Carlo simulation. arXiv:1607.03954 (2016)
41. Nüsken, N.: Construction of optimal samplers (in preparation). PhD thesis, Imperial College London (2018)
42. Ottobre, M., Pavliotis, G.A.: Asymptotic analysis for the generalized Langevin equation. Nonlinearity **24**(5), 1629 (2011)
43. Ottobre, M., Pavliotis, G.A., Pravda-Starov, K.: Exponential return to equilibrium for hypoelliptic quadratic systems. J. Funct. Anal. **262**(9), 4000–4039 (2012)
44. Ottobre, M., Pavliotis, G.A., Pravda-Starov, K.: Some remarks on degenerate hypoelliptic Ornstein-Uhlenbeck operators. J. Math. Anal. Appl. **429**(2), 676–712 (2015)
45. Ottobre, M., Pillai, N.S., Pinski, F.J., Stuart, A.M.: A function space HMC algorithm with second order Langevin diffusion limit. Bernoulli **22**(1), 60–106 (2016)
46. Pavliotis, G.A.: Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations, vol. 60. Springer, Berlin (2014)
47. Poncet, R.: Generalized and hybrid Metropolis-Hastings overdamped Langevin algorithms. arXiv:1701.05833 (2017)
48. Pavliotis, G.A., Stuart, A.M.: White noise limits for inertial particles in a random field. Multiscale Model. Simul. **1**(4), 527–533 (2003). (electronic)
49. Pavliotis, G.A., Stuart, A.M.: Analysis of white noise limits for stochastic systems with two fast relaxation times. Multiscale Model. Simul. **4**(1), 1–35 (2005). (electronic)
50. Rey-Bellet, L., Spiliopoulos, K.: Irreversible Langevin samplers and variance reduction: a large deviations approach. Nonlinearity **28**(7), 2081–2103 (2015)
51. Rey-Bellet, L., Spiliopoulos, K.: Variance reduction for irreversible Langevin samplers and diffusion on graphs. Electron. Commun. Probab., vol. 20, pp. 15, 16, (2015)
52. Robert, C., Casella, G.: Monte Carlo Statistical Methods. Springer, Berlin (2013)
53. Roussel, J., Stoltz, G.: Spectral methods for Langevin dynamics and associated error estimates. arXiv:1702.04718 (2017)
54. Villani, C.: Hypocoercivity. Number 949-951. American Mathematical Society (2009)
55. Wu, S.-J., Hwang, C.-R., Chu, M.T.: Attaining the optimal Gaussian diffusion acceleration. J. Stat. Phys. **155**(3), 571–590 (2014)