Convergence and variance reduction for stochastic differential equations in sampling and optimisation

A thesis submitted for the degree of
Doctor of Philosophy of Imperial College London
by

Martin Chak

Department of Mathematics
Imperial College London
London SW7 2AZ

September 23, 2022

I certify that the research in this thesis is the product of my own work. Any ideas or quotations from the work of others has been properly acknowledged.

Abstract

Three problems that are linked by way of motivation are addressed in this work.
In the first part of the thesis, we study the generalised Langevin equation for simulated annealing with the underlying goal of improving continuous-time dynamics for the problem of global optimisation of nonconvex functions. The main result in this part is on the convergence to the global optimum, which is shown using techniques from hypocoercivity given suitable assumptions on the nonconvex function. Alongside, we investigate numerically the problem of parameter tuning in the continuous-time equation.
In the second part of the thesis, this last problem is addressed rigorously for the underdamped Langevin dynamics. In particular, a systematic procedure for finding the optimal friction matrix in the sampling problem is presented. We give an expression for the gradient of the asymptotic variance in terms of solutions to Poisson equations and present a working algorithm for approximating its value.
Lastly, regularity of an associated semigroup, twice differentiable-in-space solutions to the Kolmogorov equation and weak numerical convergence rates of order one are shown for a class of stochastic differential equations with superlinearly growing, non-globally monotone coefficients. In the relation to the previous part, the results allow the use of Poisson equations for variations of Langevin dynamics not permissible before.

Acknowledgments

I thank my supervisor Greg Pavliotis, who guided me through scientific topics and directions with unending patience. He supported my research with enthusiasm and expertise, for which I am grateful. Thank you also to Nik Kantas, who shared his knowledge and time unreservedly. He supported me with thorough feedback and encouragement, which has played an important part in the work here.

I am especially grateful to Tony Lelièvre for our discussions. He partook in these conversations with interest and directness, which have made for satisfying experiences for me. I am also grateful for his kind hospitality during my stay at CERMICS.

There have been numerous scientific discussions with the people in the Department of Mathematics at Imperial, at conferences and workshops, at CERMICS that I should thank, amongst whom I expressly mention Urbain Vaes, Panos Parpas, Gabriel Stoltz, Rishabh Gvalani, Michele Coti Zelati, Oliver Tse, Pierre Monmarché, Gero Friesecke, Anastasia Borovykh, Daniel Lengyel and Matteo Tabaro.

Finally, thank you to my friends, those in my office and to my mother and family, who have unconditionally supported me.

Contents

List of Figures

List of Tables

1

Introduction

The purpose of this section is to present the synopsis of the thesis, to explain the connection between the problems in each of the chapters and to mention briefly the methodologies involved. Since the chapters contribute to sufficiently disjoint parts of the literature, the review of such is left to each chapter.

1.0.1 Generalised Langevin equation

The paradigmatic continuous time dynamics that is at the base of our investigations is the following. For a positive function $U \in C^1(\mathbb{R}^n)$, constant $T > 0$ and a standard Wiener process on $\mathbb{R}^n$, consider an $\mathbb{R}^n$-valued solution X_t to the stochastic differential equation (SDE)

$$dX_t = -\nabla U(X_t)dt + \sqrt{2T}dW_t,$$

that is called Langevin, Brownian or overdamped Langevin dynamics

For our purposes in sampling and optimisation problems, the interest in this model is that under suitable assumptions on U, the distribution of X_t is proportional to $\exp(-\frac{U(x)}{T})dx$ if the same holds for X_0, that is, $\exp(-\frac{U(x)}{T})dx$ is invariant. In the case of the sampling, we consider the problem itself to be drawing samples from a distribution $\exp(-\frac{U(x)}{T})dx$ for use in estimating values depending on such and in the case of optimisation, to be finding the minimum of a function U. The invariance of $\exp(-\frac{U(x)}{T})dx$ can be seen intuitively by an application of Itô's rule: for compactly supported $f \in C^2(\mathbb{R}^n)$, it holds that

$$f(X_t) = f(X_0) + \int_0^t (-\nabla U(X_s) \cdot \nabla f(X_s) + \Delta f(X_s))ds + \int_0^t \sqrt{2T}\nabla f(X_s)dW_s,$$

so that taking expectations gives

$$\mathbb{E}f(X_t) - \mathbb{E}f(X_0) = \int_0^t \mathbb{E}(-\nabla U(X_s) \cdot \nabla f(X_s) + \Delta f(X_s))ds$$

and the equation is solved by $X_s \sim \exp(-\frac{U(x)}{T})dx$ for all $s \in [0,t]$, since in that case we have

$$\begin{aligned}\int_0^t \mathbb{E}(-\nabla U(X_s)\cdot\nabla f(X_s) + \Delta f(X_s))ds &\propto \int_0^t \int_{\mathbb{R}^n} (-\nabla U(x)\cdot\nabla f(x) + \Delta f(x))e^{-U(x)}dxds \\ &= \int_0^t \int_{\mathbb{R}^n} f(x)\nabla\cdot((\nabla U(x) + \nabla)e^{-U(x)})dxds \\ &= 0.\end{aligned}$$

One may include a momentum variable in the dynamics above in order to improve the exploration of X_t over $\mathbb{R}^n$. For a constant $\gamma > 0$, the same calculation as above for

$$dX_t = Y_t dt, \tag{1.0.1}$$

$$dY_t = -\nabla U(X_t)dt - \gamma Y_t dt + \sqrt{2\gamma T}dW_t, \tag{1.0.2}$$

that is the *(underdamped) Langevin equation*, also holds but with the invariant distribution $e^{-\frac{1}{T}(U(x)+\frac{|y|^2}{2})}$, whose marginal distribution in x is the same as before and the same still for the *generalised Langevin equation*

$$\begin{aligned}dX_t &= Y_t dt, \\ dY_t &= -\nabla U(X_t)dt + \lambda Z_t dt, \\ dZ_t &= -\lambda Y_t dt - \lambda' Z_t dt + \sqrt{2\lambda' T}dW_t\end{aligned}$$

for constants $\lambda \in \mathbb{R}$, $\lambda' > 0$ with invariant distribution $e^{-\frac{1}{T}(U(x)+\frac{|y|^2}{2}+\frac{|z|^2}{2})}$.
Specifically for the problem of optimisation, one may make use of the forms of these distributions by taking $T \to 0$ as $t \to \infty$, so that the marginal distribution of X_t is concentrated around the minimum of U for large t if the distribution of X_t is indeed close to the invariant distribution. This is in the spirit of *simulated annealing*, which refers to the particular feature of gradual changes of a temperature parameter, which is T in our cases above, to a limit in order to find the global minimum value of a function. It is an open problem for general systems to show that such applications of simulated annealing indeed yield convergence to the global optimum and the rate of change in T-like parameters in the large time must typically be slow enough for the system to explore the space sufficiently. For the overdamped Langevin dynamics, convergence was shown in [42]. For the underdamped Langevin equation, an approach for this convergence was shown for the first time in [139]. The main goal in Chapter 2 is to prove that the same holds for the generalised Langevin equation. Alongside, numerical examples are given for cases of improved performance over the underdamped dynamics.

The technical core of Chapter 2 is to prove, for $T = T_t$ varying in time, m_t the density of X_t, $\mu_t(x,y,z) := e^{\frac{-1}{T_t}(U(x)+\frac{|y|^2}{2}+\frac{|z|^2}{2})}$, $Z := \int_{\mathbb{R}^n} \mu_t(x,y,z)dxdydz$, $h_t := \frac{m_t}{Z^{-1}\mu_t}$ and some constants $C > 0$, $0 < \gamma < 1$, that an inequality of the form

$$\int m_t \ln h_t \leq C\left(\frac{1}{t}\right)^{\gamma}, \tag{1.0.3}$$

holds for all $t > 0$. In order to prove such an inequality, the quantity

$$\int \left(\frac{\langle S\nabla h_t, \nabla h_t\rangle}{h_t} + \beta(T_t^{-1})h_t \ln h_t\right) Z^{-1} d\mu_t,$$

for some well chosen matrix S and polynomial $\beta(\cdot)$ will be considered. This construction of H over the left-hand side of (1.0.3) is the strategy of [190] and, as is well-known, compensates for the fact that the diffusion is degenerate.

1.0.2 Formula for improving a certain parameter

The analysis described above turns out not to guide the choice of parameter in the generalised Langevin equation. We are motivated then in the next part of the thesis to find a methodology for choosing parameters in the dynamics. In this direction, focus is placed on the somewhat simpler underdamped Langevin dynamics (1.0.1) and the friction parameter γ for the sampling problem. It is worth highlighting that for this problem, although the goal is to find optimal parameters for the dynamics, the sense of optimality is not a speed of convergence to a global optimum of U, but to be decided out of a number of options relevant for the sampling problem. In particular, our analysis is aimed at reducing the value of the *asymptotic variance* for a function $f : \mathbb{R}^{2n} \to \mathbb{R}^n$ given by the expression

$$\sigma^2 := 2\int \phi(f - \int f d\pi) d\pi, \tag{1.0.4}$$

where ϕ is the solution to

$$L\phi(x,y) = f(x,y) - \int f d\pi, \tag{1.0.5}$$

the operator L is the infinitesimal generator of the underdamped Langevin dynamics given by

$$L = -y \cdot \nabla_x + \nabla U(x) \cdot \nabla_y + \gamma y \cdot \nabla_y - \gamma \nabla_y \cdot \nabla_y \tag{1.0.6}$$

when acting on smooth functions and π is the invariant measure with density proportional to $e^{-U(x)-\frac{|y|^2}{2}}$. The use of (1.0.4) and (1.0.5) are justified as follows. For solutions (X_t, Y_t)

to the underdamped Langevin equation, the random variable

$$\frac{1}{\sqrt{t}}\int_0^t \left(f(X_s,Y_s)-\int f d\pi\right)ds, \tag{1.0.7}$$

converges in distribution to $\mathcal{N}(0,\sigma^2)$ under suitable assumptions on U and f, so that σ determines an asymptotic speed for the convergence of the time average $\frac{1}{t}\int_0^t f(X_s,Y_s)ds$ to a quantity of interest $\int f d\pi$, the approximation of which we assume to be the sampling problem. The next paragraph is devoted to a formal justification for this convergence, see the functional central limit theorem approach in [15] and also [113]. Supposing a solution ϕ to (1.0.5) indeed exists and that it is smooth, by Itô's rule we have

$$\phi(X_t,Y_t)=\phi(X_0,Y_0)-\int_0^t (f(X_s,Y_s)-\int f d\pi)ds+\sqrt{2\gamma}\int_0^t \nabla_y\phi(X_s,Y_s)\cdot dW_s, \tag{1.0.8}$$

so that assuming initial stationarity $(X_0,Y_0)\sim\pi$ and taking expectation gives

$$\mathbb{E}\left[\frac{1}{\sqrt{t}}\int_0^t (f(X_s,Y_s)-\int f d\pi)ds\right]=0.$$

Moreover, using (1.0.8), the variance of (1.0.7) can be calculated as

$$\begin{aligned}
&\frac{1}{t}\mathbb{E}\bigg[(\phi(X_t,Y_t)-\phi(X_0,Y_0))^2-2\sqrt{2\gamma}(\phi(X_t,Y_t)-\phi(X_0,Y_0))\bigg(\int_0^t \nabla_y\phi(X_s,Y_s)\cdot dW_s\bigg)\\
&\quad+2\gamma\bigg(\int_0^t \nabla_y\phi(X_s,Y_s)\cdot dW_s\bigg)^2\ \bigg]\\
&\leq\frac{2}{t}\mathbb{E}\bigg[(\phi(X_t,Y_t)-\phi(X_0,Y_0))^2+2\gamma\bigg(\int_0^t \nabla_y\phi(X_s,Y_s)\cdot dW_s\bigg)^2\bigg]
\end{aligned}$$

Given that $t^{-\frac{1}{2}}\phi(X_t,Y_t)$ converges to zero in square mean as $t\to\infty$ (it will be shown that in fact equation (1.0.10) holds), the only nonzero term as $t\to\infty$ is the last term, for which formally we have

$$\begin{aligned}
2\gamma\mathbb{E}\bigg[\bigg(\frac{1}{t}\int_0^t \nabla_y\phi(X_s,Y_s)\cdot dW_s\bigg)^2\bigg]&=\frac{2\gamma}{t}\int_0^t \nabla_y\phi(X_s,Y_s)\cdot\nabla_y\phi(X_s,Y_s)ds\\
&\to 2\gamma\int \nabla_y\phi\cdot\nabla_y\phi\, d\pi\\
&=-2\gamma\int \phi(\nabla_y\cdot\nabla_y-y\cdot\nabla_y)\phi)d\pi\\
&=-2\int \phi\bigg(f-\int f d\pi\bigg)d\pi
\end{aligned}$$

as $t \to \infty$, where Itô's isometry, (assumed) ergodicity, (1.0.5), (1.0.6) and integration by parts have been used.

Having justified the validity of the criterion of asymptotic variance and returning to its representation (1.0.4), the main result of Chapter 3 is a formula for the derivative of the asymptotic variance with respect to γ. In particular, it is shown that

$$\frac{\partial}{\partial\gamma}\sigma^2 = -2\int \nabla_y\phi \cdot \nabla_y\tilde{\phi}d\pi, \tag{1.0.9}$$

where $\tilde{\phi}(x,y) = \phi(x,-y)$. Therefore given approximations of $\nabla_y\phi$, equation (1.0.9) naturally yields a gradient descent procedure on σ^2 with respect to γ. We may indeed approximate $\nabla_y\phi$ by using the known formula

$$\phi(x,y) = \int_0^\infty (\mathbb{E}f(X_t,Y_t) - \int f d\pi)dt, \tag{1.0.10}$$

where (X_t, Y_t) is a solution to the underdamped Langevin equation with initial condition (x,y). In particular, one has a systematic procedure for improving the value of γ for the goal of approximating $\int f d\pi$ using $\int_0^t f(X_s,Y_s)ds$. It is shown in Chapter 3 that this provides a functioning gradient procedure for estimating improved values of γ.

1.0.3 Kolmogorov equations

In order to derive (1.0.9), we make use of the fact that the differential operator (1.0.6) has a maximally accretive closure in $L^2(\pi)$, see Section 5.2 in [91] for a definition and discussion. Existing results about the maximal accretivity of the closure of analogous operators to the above are far from general when noise does not appear in all of the component dynamics of the SDE. An alternative, more involved method for proving the main formula (1.0.9) of Chapter 3 (not explicitly given here, but see [36]) is to use that an associated semigroup is the solution to a Kolmogorov equation.

In Chapter 4, we consider general SDEs driven by Wiener processes. Our underlying result is to show, by combining techniques of [44] with those in [114], some moment estimates on derivative (with respect to initial condition) processes for SDEs with nonglobally monotone coefficients. These estimates are bedrock to three results that give

- differentiability-in-x of the semigroup $P_tg(x) := \mathbb{E}g(X_t)$ for g regular enough, where X_t solves the SDE with initial condition x,
- that the semigroup solves the Kolmogorov equation $(\partial_t + \mathcal{L})P_tg = 0$, where $\mathcal{L}$ is the generator of the SDE acting on twice differentiable functions and

- weak numerical convergence rates of order one for a stopped increment-tamed Euler-Maruyama scheme.

In the case where the coefficients of the SDE are globally Lipschitz or at least globally monotone, such results are known and in fact numerous related works are available on numerical convergence rates. However, Hairer et al. [85] showed that there exist counterexamples to the aforementioned differentiability and weak polynomial convergence rates for the Euler-Maruyama scheme outside this regime even in the case where the coefficients are smooth and globally bounded. For example, from Theorem 3.1 in [85], solutions up to time $T > 0$ to the SDE

$$\begin{aligned} dX_t &= \cos(Z_t \cdot \exp(Y_t^3))dt \\ dY_t &= \sqrt{2}dW_t \\ dZ_t &= 0dt \end{aligned}$$

are such that there exist $\varphi \in C_c^\infty$ where $\mathbb{E}[\varphi(X_t, Y_t, Z_t)]$ is not locally Hölder continuous in initial condition for any $t \in (0, T]$. Our main assumption here is loosely that the Lipschitz constant of our drift and diffusions coefficients are $o(\log V)$ and $o(\sqrt{\log V})$ respectively for a Lyapunov function V satisfying $LV \leq CV$, for some constant C and the associated differential operator L (for example given by (1.0.6) for the SDE (1.0.1)-(1.0.2)). Our conditions are satisfied by a class of SDEs similar to [44, 105] and show for the first time all of the above under these conditions.
In particular, these results allow one to show that the semigroup associated to the underdamped Langevin equation with variable friction $\gamma = \gamma(x, y)$, which has non-globally Lipschitz coefficients by definition, solves the Kolmogorov equation and as a consequence forms a distributional solution to the equation (1.0.5), thus forming a first step for proving formulae such as (1.0.9).

2

Generalised Langevin equation for simulated annealing

The contents of this chapter are from the paper [37] written in collarboration with G. Pavliotis and N. Kantas.

2.1 Introduction

Optimisation algorithms have received significant interest in recent years due to applications in machine learning, data science and molecular dynamics. Many models in machine learning result in a formulation whereby some loss function and its parameters are to be minimised, in which use of optimisation techniques is heavily relied upon. We refer to [22, 179] for related discussions. Many models, for instance neural networks, use parameters that vary over a continuous space, where gradient-based optimisation methods can be used to find good parameters that generate effective predictive ability which fulfill the purpose of the model. As such, the design and analysis of such algorithms for global optimisation has been the subject of considerable research [170] and it has proved useful to study algorithms for global optimisation using tools from the theory of stochastic processes and dynamical systems. A paradigm of the use of stochastic dynamics for the design of algorithms for global optimisation is one of simulated annealing, where overdamped Langevin dynamics with a time dependent temperature (2.1.1) that decreases with an appropriate cooling schedule is used to guarantee the global minimum of a nonconvex loss function $U : \mathbb{R}^n \to \mathbb{R}$:

$$dX_t = -\nabla U(X_t)\, dt + \sqrt{2T_t}\, dW_t. \tag{2.1.1}$$

Here W_t is a standard n-dimensional Wiener process and $T_\cdot : (0, \infty) \to (0, \infty)$ is an appropriate determinstic function of time often referred to as the annealing or cooling schedule.

For fixed $T_t = T > 0$, this is the dynamics used for the related problem of sampling from a possibly high dimensional probability measure, for example in the unadjusted Langevin algorithm [59]. Gradually decreasing T_t to zero balances the exploration-exploitation trade-off by allowing at early times larger noise to drive X_t and hence sufficient mixing to escape local minima. Designing an appropriate annealing schedule is well-understood. We briefly mention classical references [42, 73, 74, 77, 78, 92, 93, 115], as well as the more recent [107, 131, 160], where one can find details and convergence results. In this chapter we aim to consider generalised versions of (2.1.1) for the same purpose.
Using dynamics such as (2.1.1) has clear connections with sampling, as stated in the introduction of the thesis. When $T_t = T$ is a constant function, the invariant distribution of X is proportional to $\exp(-\frac{U(x)}{T})dx$. In addition, when T_t decreases with time, the probability measure given by $\nu_t(dx) \propto \exp(-\frac{U(x)}{T_t})dx$ converges weakly to the set of global minima based on the Laplace principle [106]. One can expect that if one replaces (2.1.1) with a stochastic process that mixes faster and maintains the same invariant distribution for constant temperatures, then the superior speed of convergence should improve performance in optimisation due to the increased exploration of the state space. Indeed, it is well known that many different dynamics can be used in order to sample from a given probability distribution, or for finding the minima of a function when the dynamics is combined with an appropriate cooling schedule for the temperature. Different kinds of dynamics have already been considered for sampling, e.g. nonreversible dynamics, preconditioned unadjusted Langevin dynamics [3, 10, 125, 156], as well as for optimisation, e.g. interacting Langevin dynamics [184], consensus based optimisation [27, 28, 161], to name a few.
A natural candidate in this direction is to use the underdamped Langevin dynamics:

$$dX_t = Y_t\,dt \tag{2.1.2a}$$
$$dY_t = -\nabla U(X_t)\,dt - T_t^{-1}\mu Y_t\,dt + \sqrt{2\mu}\,dW_t. \tag{2.1.2b}$$

Here the reversibility property of (2.1.1) has been lost; the improvement from breaking reversibility in both the context of sampling and that of optimisation is investigated in [55, 122] and [70] respectively. When $T_t = T$, (2.1.2) can converge faster than (2.1.1) to its invariant distribution

$$\rho(dx, dy) \propto \exp\left(-\frac{1}{T}\left(U(x) + \frac{|y|^2}{2}\right)\right)dx\,dy,$$

see [61] or Section 6.3 of [158] for particular (theoretical) comparisons and also [18, 19] for more applications using variants of (2.1.2). In the context of simulated annealing, using this set of dynamics has recently been studied rigorously in [139], where the author established convergence to global minima using the generalised Γ-calculus [140] framework that generalises Bakry-Emery theory to degenerate systems. Note that (2.1.2b) uses the temperature in the drift rather than the diffusion constant in the noise as in (2.1.1). Both formulations admit the same invariant measure when $T_t = T$. In the remainder of the chapter, we adopt this formulation to be closer to [139].

In this chapter we will consider an extension of the kinetic Langevin equation by adding an additional auxiliary variable that accounts for the memory in the system. To the best of the authors' knowledge, this has not been attempted before in the context of simulated annealing and global optimisation. In particular we consider the Markovian approximation [158, Section 8.2] to the generalised Langevin equation:

$$dX_t = Y_t\, dt \tag{2.1.3a}$$

$$dY_t = -\nabla U(X_t)\, dt + \lambda^\top Z_t\, dt \tag{2.1.3b}$$

$$dZ_t = -\lambda Y_t\, dt - T_t^{-1} A Z_t\, dt + \Sigma\, dW_t, \tag{2.1.3c}$$

where $A \in \mathbb{R}^{m\times m}$ is symmetric positive definite matrix, meaning that there exists a constant $A_c > 0$ such that $z^\top A z \geq A_c |z|^2$ for all $z \in \mathbb{R}^m$, $\Sigma \in \mathbb{R}^{m\times m}$ satisfies $\Sigma\Sigma^\top = 2A$ (that is, the fluctuation-dissipation theorem [158, Section 6.1] holds) and W_t is now m-dimensional. Here $X_t, Y_t \in \mathbb{R}^n$ and $Z_t \in \mathbb{R}^m$ (with $m \geq n$), $M^\top$ denotes the transpose of a matrix M, $\lambda \in \mathbb{R}^{m\times n}$ is a rank n matrix with a left inverse $\lambda^{-1} \in \mathbb{R}^{n\times m}$.

Our aim is to establish convergence using similar techniques as [139] and investigate the improvements in performance. Equation (2.1.3) is related to the generalised Langevin equation, where memory is added to (2.1.2) by integrating over past velocities with a kernel $\Gamma : (0,\infty) \to \mathbb{R}^{n\times n}$:

$$\ddot{x} = -\nabla U(x) - \int_0^t \Gamma(t-s)\dot{x}(s)\, ds + F_t \tag{2.1.4}$$

with F_t being a zero mean stationary Gaussian process with an autocorrelation matrix given by the fluctuation-dissipation theorem $\mathbb{E}(F_t F_s^\top) = T_t \Gamma(t-s)$. When[1] $T_t = T$, (2.1.4) is equivalent to (2.1.3) with $Z_0 \sim \mathcal{N}(0, TI)$ for identity matrix I when setting $\Gamma(t) = \lambda^\top e^{-At}\lambda$, see Proposition 8.1 in [158]. In this case, the invariant distribution becomes

$$\rho(dx, dy, dz) \propto \exp\left(-\frac{1}{T}\left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right)\right) dx\, dy\, dz.$$

In the spirit of adding a momentum variable in (2.1.1) to get (2.1.2), (2.1.3) adds an additional auxiliary variable to the Langevin system whilst preserving the invariant dis-

[1] To our knowledge, there is no known direct translation between (2.1.4) and (2.1.3) for a non-constant T_t; at the very least the intuition here is useful.

tribution in the x marginal. In the constant temperature context, (2.1.4) is natural from the point of view of statistical mechanics and has already been considered as a sampling tool in [31, 32, 33, 145] with considerable success. We will demonstrate numerically that the additional tuning parameters can improve performance; see also [143] for recent work demonstrating advantages of using (2.1.4) compared to using (2.1.2) when sampling from a log concave density. A detailed study of the Markovian approximation (2.1.3) of the generalised Langevin dynamics in (2.1.4) can be found in [151].

To motivate the use of (2.1.3), consider the quadratic case where $U = \alpha x^2$ and $0 < \alpha < 1$. This case allows for explicit or numerical calculation of the spectral gaps of the generators in (2.1.1)-(2.1.3) in order to compare the rate of convergence to equilibrium; see [135, 152] for details. Preliminary work not explicitly presented here show that for the aforementioned cases, best choices of λ, A yield an improvement in terms of the spectral gap compared to (2.1.2) with the best choice of μ.

Use of (2.1.4) is also motivated by parallels with accelerated gradient descent algorithms. When the noise is removed from (2.1.2), the second order differential equation can be loosely considered as a continuous time version of Nesterov's algorithm [183]. The latter is commonly preferred to discretising the first order differential equation given by the noiseless version of (2.1.1), because in the high dimensional and low iterations setting it achieves the optimal rate of convergence for convex optimisation; see Chapter 2 in [150] and also [76] for a nonconvex setting. Here we would like to investigate the effect of adding another auxiliary variable, which would correspond to a third order differential equation when noise is removed. When noise is added for the fixed temperature case, [69] has studied the long time behaviour and stability for different choices of a memory kernel as in (2.1.4). Finally, we note that generalised Langevin dynamics in (2.1.4) have additionally been studied in related areas such as sampling problems in molecular dynamics from chemical modelling [1, 31, 32, 33, 145, 195], see also [116] for work determining the kernel Γ in the generalised system (2.1.4) from data.

Our theoretical results will focus only on the continuous time dynamics and follow the approach in [139]. The main requirement in terms of assumptions are quadratic upper and lower bounds on U and bounded second derivatives. This is different to classical references such as [74], [77] or [93]. These works also rely on the Poincaré inequality, an approach which will be mirrored here (and in [139] for the underdamped case) using a log-Sobolev inequality; see also [92] for the relationship between such functional inequalities and the annealing schedule in the finite state space case. We will also present detailed numerical results for different choices of U. There are many possibilities for the method of discretisation of (2.1.3), we will use a time discretisation scheme that appeared in [6], but will not present theoretical results on the time discretised dynamics; this is beyond the scope of this thesis. We refer instead the interested reader to [171] for a study on discretisation schemes for the system (2.1.3), [41] for a recent consideration on (2.1.2) and its time-discretisation and [71, 72] for linking discrete time Markov chains with the overdamped Langevin system in (2.1.1).

2.1.1 Contributions and organisation of the chapter

Here we summarise the main contributions of the chapter.

- We provide a complete theoretical analysis of the simulated annealing algorithm for the generalised Langevin equation (2.1.3). The main theoretical contribution consists of Theorem 2.2.5 that establishes convergence in probability of X_t in the higher order Markovian dynamics (2.1.3) to a global minimiser of U. For the optimal cooling schedule T_t out of those that are proved here to give a convergent process, the rate of convergence is as the known rate for the Langevin system (2.1.2) presented in [139].

- The initially non-Markovian property and pronounced degeneracy in the sense of requiring a second commutator bracket for hypoellipticity by way of Hörmander introduces additional difficulties that are overcome using techniques from [139]. As such, we use a different form of the distorted entropy, stated formally in (2.4.36). Additional technical improvements include a different truncation argument and a limiting sequence of nondegenerate SDEs for establishing dissipation of this distorted entropy. These extensions also address certain technical issues in [139]; see Remarks 2.2.1, 2.4.1 and 2.4.4 for more details. Also we make an effort to emphasise the role of the critical factor of the cooling schedule in the rate of convergence in Theorem 2.2.5. This can be seen in our assumptions for T_t and U below.

- Numerical experiments are provided to illustrate the performance of our approach. We also discuss tuning issues. In particular, we investigate numerically the role of matrix A and how it can be chosen to increase exploration of the state space. As regards to time discretisation of (2.1.3) we use the leapfrog scheme of [6]. We compare this with a similar time discretisation of (2.1.2) and observe that exploration of the state space is increased considerably.

The rest of the chapter is organised as follows. Section 2.2 will present the assumptions and main theoretical results. Proofs can be found in Section 3.6. Section 2.3 presents numerical results demonstrating the effectiveness of our approach in terms of reaching the global minimum. In Section 2.6, we provide some concluding remarks.

2.2 Main Result

Let L_t denote the infinitesimal generator of the associated semigroup to (2.1.3) at $t > 0$ and temperature T_t. This is formally given by

$$L_t = (y \cdot \nabla_x - \nabla_x U(x) \cdot \nabla_y) + (z^\top \lambda \nabla_y - y^\top \lambda^\top \nabla_z) - T_t^{-1} z^\top A \nabla_z + A : D_z^2, \qquad (2.2.1)$$

where we denote the gradient vector as $\nabla_x = (\partial_{x_1}, \dots, \partial_{x_n})^\top$, the Hessian with D_x^2 and similarly for the y and z variables. For matrices $M, N \in \mathbb{R}^{r\times r}$ we denote $M : N = \sum_{i,j} M_{ij} N_{ij}$ for all $1 \le i, j \le r$ and the operator norm as

$$|M| = \sup\left\{ \frac{|Mv|}{|v|} : v \in \mathbb{R}^r \text{ with } v \neq 0 \right\}.$$

We will also use $|v|$ to denote Euclidean distance for a vector v. Let m_t be the law of (X_t, Y_t, Z_t) in (2.1.3) and, with slight abuse of notation, we will also denote as m_t the corresponding Lebesgue density. Similarly we define μ_{T_t} be the instantaneous invariant law of the process

$$\mu_{T_t}(dx, dy, dz) \quad = \quad \frac{1}{Z_{T_t}} \exp\Bigg(- \frac{1}{T_t}\bigg(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \bigg)\Bigg) dx\, dy\, dz \qquad (2.2.2)$$

with $Z_{T_t} = \int \exp\big(- \frac{1}{T_t}\big(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2}\big)\big) dx\, dy\, dz$. Finally, denote the density between the two laws $h_t = \frac{dm_t}{d\mu_{T_t}}$. We proceed by stating our assumptions on the potential U.

Assumption 1. The function U belongs to $\mathcal{C}^\infty(\mathbb{R}^n)$ and its second derivatives satisfy

$$|D_x^2 U|_\infty := \sup_{x\in\mathbb{R}^n} \max\left\{ \sup_{ij} |\partial_i \partial_j U(x)|, |D_x^2 U(x)| \right\} < \infty. \qquad (2.2.3)$$

Its first derivatives satisfy

$$\nabla_x U(x) \cdot x \ge r_1 |x|^2 - U_g \qquad (2.2.4)$$

$$|\nabla_x U(x)|^2 \le r_2 |x|^2 + U_g \qquad (2.2.5)$$

for some constants $r_1, r_2 \in \mathbb{R}$, $U_g > 0$. Moreover, either

(a)

$$|\bar{a} \circ x|^2 + U_m \quad \le \quad U(x) \quad \le \quad |\bar{a} \circ x|^2 + U_M \qquad (2.2.6)$$

for some $U_m, U_M \in \mathbb{R}$, $\bar{a} \in (0, \infty)^n$, where $\circ$ denotes the Hadamard product, or

(b) • U is a nonnegative Morse function, defined as follows. There exists $1 \le C_H <$

∞ such that if $x \in \mathbb{R}^n$ satisfies $\nabla_x U(x) = 0$, then

$$\frac{1}{C_H} \leq \|D_x^2 U(x)\| \leq C_H$$

- U is nondegenerate in the sense that:
  - For any two local minima $m_i, m_j \in \mathbb{R}^n$, there exists a unique (communicating saddle) point $s_{i,j} \in \mathbb{R}^n$ such that
    - $\nabla_x U(s_{i,j}) = 0$,
    - $U(s_{i,j}) = \inf\{\max_{s\in[0,1]} U(\gamma(s)) : \gamma \in C([0,1], \mathbb{R}^n), \gamma(0) = m_i, \gamma(1) = m_j\}$,
    - the dimension of the unstable subspace of $D_x^2 U(s_{i,j})$ is equal to 1.
  - Setting m_1 to be the global minimum of U, there exists $\delta > 0$ and an ordering of the local minima $\{m_2, m_3, \dots\}$ such that $U(s_{1,2}) - U(m_2) \geq U(s_{1,i}) - U(m_i) + \delta$ for all $i \geq 3$.

Note that (2.2.4) and (2.2.5) imply

$$a_m|x|^2 + U_m \quad \leq \quad U(x) \quad \leq \quad a_M|x|^2 + U_M \tag{2.2.7}$$

for some $a_m, a_M > 0$, $U_m, U_M \in \mathbb{R}$. In the rest of the chapter, if (2.2.6) holds then the smallest and largest element of $\bar{a}$ is denoted with $a_m = \min_i \bar{a}_i$ and $a_M = \max_i \bar{a}_i$, where $\bar{a} = (\bar{a}_1, \dots, \bar{a}_n)$.

Assumption 2. The temperature T_t satisfies $\lim_{t\to\infty} T_t = 0$.

Before we proceed with further assumptions on the annealing schedule T_t and on the initial distribution, note that under Assumption 1 and 2, a log-Sobolev inequality holds with a time-varying constant that increases as $t \to \infty$, which is consistent with the concentration of μ_{T_t} around global minima of U. This allows one to conclude exponential convergence to an instantaneous equilibrium at each t and forms part of the proof to our main convergence result.

Proposition 2.2.1. *Under Assumptions 1 and 2, there exist constants $t_{ls}^{(0)}$, $\hat{E}$, $A_*^{(0)} > 0$ and a finite order polynomial $r^{(0)} : (0,\infty) \to (0,\infty)$ with coefficients depending on U such that for all $0 < h \in C^\infty(\mathbb{R}^{2n+m})$ satisfying $\int h d\mu_{T_t} = 1$, it holds that*

$$\int h \ln h d\mu_{T_t} \leq C_t^{(0)} \int \frac{|\nabla h|^2}{h} d\mu_{T_t}, \tag{2.2.8}$$

where for $t > t_{ls}^{(0)}$,

$$C_t^{(0)} = r^{(0)}\left(T_t^{-\frac{1}{2}}\right)e^{\hat{E}T_t^{-1}}. \tag{2.2.9}$$

Proof. Firstly, the case that U satisfies Assumption 1(a) is dealt with. The standard log-Sobolev inequality for a Gaussian measure [80] alongside the properties that log-Sobolev inequalities tensorises and are stable under perturbations, which can be found as Theorem 4.4 and Property 4.6 in [83] respectively, yields the result. In particular,

$$\begin{aligned}
\int h\ln h d\mu_{T_t} &= \int (h\ln h - h + 1)d\mu_{T_t}\\
&\leq \int (h\ln h - h + 1)Z_{T_t}^{-1}e^{-\frac{U_m}{T_t}-\frac{1}{T_t}\left(|\bar{a}\circ x|^2+\frac{|y|^2}{2}+\frac{|z|^2}{2}\right)}dxdydz\\
&= e^{-\frac{U_m}{T_t}}Z_{T_t}^{-1}\int h\ln h e^{-\frac{1}{T_t}\left(|\bar{a}\circ x|^2+\frac{|y|^2}{2}+\frac{|z|^2}{2}\right)}dxdydz\\
&\leq e^{-\frac{U_m}{T_t}}\max\left(\frac{T_t}{2},\max_i\frac{T_t}{4\bar{a}_i^2}\right)Z_{T_t}^{-1}\int\frac{|\nabla h|^2}{h_t}e^{-\frac{1}{T_t}\left(|\bar{a}\circ x|^2+\frac{|y|^2}{2}+\frac{|z|^2}{2}\right)}dxdydz\\
&\leq e^{\frac{U_M-U_m}{T_t}}\max\left(\frac{T_t}{2},\frac{T_t}{4a_m^2}\right)\int\frac{|\nabla h|^2}{h}d\mu_{T_t},
\end{aligned}$$

where the first inequality follows by (2.2.6) since $x\ln x - x + 1 \geq 0$ for all $x \geq 0$, so that

$$C_t^{(0)} = \max\left(2, a_m^{-2}\right)\frac{T_t}{4}e^{(U_M-U_m)T_t^{-1}}.$$

In the case of Assumption 1(b), the inequality in the x-marginals is taken as a consequence of Corollary 2.17 in [134] (see however Definition 1.8 in [176] for Morse functions); for the announced form (2.2.9) of $C_t^{(0)}$, equation (2.18) in [134] can be used by taking $t_{ls}^{(0)}$ large enough such that for $t > t_{ls}^{(0)}$, T_t is small enough. The proof concludes by [83, Theorem 4.4] together with the log-Sobolev inequality for Gaussian measures. ☐

The constant $\hat{E}$ from the above proposition will be used in stating the following assumption about T_t, as well as what follows. In the case of Assumption 1(a), $\hat{E}$ can be taken as $U_M - U_m$, otherwise for Assumption 1(b) it is the critical depth [134] of U.

Assumption 3. The cooling schedule $T_\cdot : [0,\infty) \to (0,\infty)$ is continuously differentiable, bounded above and there exists some constant $t_0 > 1$ such that T_t satisfies for all $t > t_0$:

(i) $T_t \geq E(\ln t)^{-1}$ for some constant $E > \hat{E} \geq 0$, where $\hat{E}$ is the constant in Proposition 2.2.1,

(ii) Denoting $T_t' = \partial_s T_s|_{s=t}$, it holds that $|T_t'| \leq \widetilde{T}t^{-1}$ for some constant $\widetilde{T} > 0$.

Assumption 4. The initial law m_0 admits a bounded density with respect to the Lebesgue measure on $\mathbb{R}^{2n+m}$, also denoted m_0, satisfying:

(i) $m_0 \in \mathcal{C}^\infty(\mathbb{R}^{2n+m})$,

(ii) $\int \frac{|\nabla m_0|^2}{m_0} dxdydz < \infty$,

(iii) $\int (|x|^2 + |y|^2 + |z|^2) m_0 \, dxdydz < \infty$,

Remark 2.2.1. Note that (2.2.5) and (2.2.6) deviate from [139]. Condition (2.2.6) is useful for a self-contained exposition for the log-Sobolev constant in (2.4.45); it is satisfied for instance by a multivariate Gaussian after a rotation of the x coordinates. The alternative condition that U is a nondegenerate Morse function allows us to conveniently apply the results of [134], in which case $\hat{E}$ is given as the critical depth of U.

We present two key propositions.

Proposition 2.2.2. *Under Assumptions 1 and 3, for all $t > 0$, denote by $\left(X^{T_t}, Y^{T_t}, Z^{T_t}\right)$ a r.v. with distribution μ_{T_t}. For any $\delta, \alpha > 0$, there exists a constant $\hat{A} > 0$ such that*

$$\mathbb{P}\big(U\big(X^{T_t}\big) > \min U + \delta\big) \leq \hat{A} e^{-\frac{\delta - \alpha}{T_t}} \tag{2.2.10}$$

holds for all $t > 0$.

Proof. The result follows exactly as in Lemma 3 in [139]. ☐

In fact, the $e^{\frac{\alpha}{T_t}}$ factor on the right-hand side in (2.2.10) may be substituted by a subexponential (in T_t^{-1}) term, see the comment after Lemma 3 in [139].

Proposition 2.2.3. *Under Assumptions 1, 3 and 4, for all $t > 0$, (X_t, Y_t, Z_t) are well defined as the unique strong solution to* (2.1.3), $\mathbb{E}\left[|X_t|^2 + |Y_t|^2 + |Z_t|^2\right] < \infty$ *and the law m_t admits an everywhere positive density with respect to the Lebesgue measure on $\mathbb{R}^{2n+m}$.*

For the proof of Proposition 2.2.3, see Proposition 2.4.1.
Proposition 2.2.2 can be thought of as a Laplace principle; Proposition 2.2.3 asserts that the process (2.1.3) does not blow up in finite time and the noise in the dynamics (2.1.3c) for Z_t spreads throughout the system, that is to X_t and Y_t.

Proposition 2.2.4. *Under Assumption 1, 3 and 4, for any $0 < \alpha \leq \frac{1}{2} - \frac{\hat{E}}{2E}$, there exists some constant $B > 0$ such that for all $t \geq 0$,*

$$\int h_t \ln h_t d\mu_{T_t} \leq B \left(\frac{1}{t}\right)^{1 - \frac{\hat{E}}{E} - 2\alpha}. \tag{2.2.11}$$

The full proof is contained in Section 2.4.7 and follows from Proposition 2.4.10. It uses an approximating sequence of SDE's, in which all of the elements have nondegenerate noise. The problem is split into the partial time and partial temperature derivatives where, amongst other tools, (2.4.40) and a log-Sobolev inequality are used as in [139] to arrive at a bound that allows a Grönwall-type argument.

Remark 2.2.2. Proposition 2.4.10 is a statement about the distorted entropy $H(t)$, which bounds the entropy $\int h_t \ln h_t d\mu_{T_t}$. In fact this is achieved in such a way that the bound becomes less sharp as t becomes large but without consequences for our main Theorem 2.2.5 below.

We proceed with the statement of our main result, using t_h from Proposition 2.2.4.

Theorem 2.2.5. *Under Assumptions 1, 2, 3 and 4, for any $\delta > 0$, as $t \to \infty$,*

$$\mathbb{P}(U(X_t) \leq \min U + \delta) \to 1.$$

If in addition $T_t = E(\ln t)^{-1}$, then for any $0 < \alpha \leq \min\left(\frac{1}{2} - \frac{\hat{E}}{2E}, \delta\right)$, there exists a constant $C > 0$ such that for all $t \geq 0$,

$$\mathbb{P}(U(X_t) > \min U + \delta) \leq C\left(\frac{1}{t}\right)^{r^e(E)},$$

where the rate $r^e : (\hat{E}, \infty) \to \mathbb{R}$ is defined by

$$\begin{aligned} r^e(E) &:= \min\left(\frac{1 - \frac{\hat{E}}{E} - 2\alpha}{2}, \frac{\delta - \alpha}{E}\right) \\ &= \begin{cases} \frac{1}{2}\left(1 - \frac{\hat{E}}{E} - 2\alpha\right) & \text{if } E < \frac{\hat{E} + 2(\delta - \alpha)}{1 - 2\alpha} \\ \frac{\delta - \alpha}{E} & \text{otherwise.} \end{cases} \end{aligned}$$

Proof. For all $t > 0$, denote by $\left(X^{T_t}, Y^{T_t}, Z^{T_t}\right)$ a random variable with distribution μ_{T_t}. For all $\delta > 0$, with the definition of h_t and triangle inequality, we have

$$\mathbb{P}(U(X_t) > \min U + \delta) \leq \mathbb{P}\big(U\big(X^{T_t}\big) > \min U + \delta\big) + \int |h_t - 1| d\mu_{T_t}.$$

Pinsker's inequality gives

$$\int |h_t - 1| d\mu_{T_t} \leq \left(2\int h_t \ln h_t d\mu_{T_t}\right)^{\frac{1}{2}}, \tag{2.2.12}$$

which, by Proposition 2.2.4, together with Proposition 2.2.2 gives the result. □

The cooling schedule $T_t = E(\ln t)^{-1}$ is optimal with respect to the method of proof for Proposition 2.4.10; see Proposition 2.5.2. This is a consistent with works in simulated annealing, e.g. [42, 73, 74, 77, 78, 92, 93, 115].

The 'mountain-like' shape of r^e indicates the bottleneck for the rate of convergence at low and high values of E: a small E means the term related to convergence to the measures μ_{T_t} is restrictive and a large E means the convergence of μ_{T_t} to the global minima of U is slow.

Although the focus in Theorem 2.2.5 is for decaying T_t, it is only for convergence to the global minimum where Assumption 2 is used. In particular, the convergence result in Proposition 2.2.4 is valid for temperature schedules that are not converging to zero. This includes the instance of using a variable temperature in order to tackle the problem of metastability in the sampling problem.

2.3 Numerical results

Here we investigate the numerical performance of (2.1.3) in terms of convergence to a global optimum and exploration capabilities and compare with (2.1.2). The details of the discretisations we use for both sets of dynamics and some details related to the annealing schedule and parameters can be found in Section 2.3.1. Rates of transition between different regions of the state space are presented in Section 2.3.2. In Section 2.3.3, for different parameters and cost functions, we present results for the probability of convergence to the global minimum. We investigate the effect of E appearing in the annealing schedule as well as the parameters in the dynamics (2.1.2) and (2.1.3). In particular, we consider different $\lambda = \bar{\lambda}\lambda_i$ and $A = \mu A_i$ in the generalised Langevin dynamics for $\bar{\lambda}, \mu > 0$. Note that μ is used also as the friction parameter in (2.1.2), which makes notational sense because μ determines the relative strength of Ornstein-Uhlenbeck part of the respective dynamics. In addition, we introduce a coefficient $\gamma > 0$ in front of the terms in (2.1.2) and (2.1.3) corresponding to the part in the respective generators given by $y \cdot \nabla_x - \nabla_x U(x) \cdot \nabla_y$; unless otherwise stated, we keep $\gamma = 1$. The numerical scheme for the experiments is detailed, but error analysis related to the discretisation is outside the scope of the thesis.

2.3.1 Time discretisation

In order to simulate from (2.1.3), we will use the following time discretisation. For $k \in \mathbb{N}$,

$$Y_{k+\frac{1}{2}} = Y_k - \frac{\Delta t}{2}\gamma\nabla U(X_k) + \frac{\Delta t}{2}\lambda^\top Z_k, \tag{2.3.1a}$$

$$X_{k+1} = X_k + \Delta t\gamma Y_{k+\frac{1}{2}}, \tag{2.3.1b}$$

$$Z_{k+1} = Z_k - \theta\lambda Y_{k+\frac{1}{2}} - \theta A Z_k + \alpha\sqrt{T_k}\ \Sigma\xi_k, \tag{2.3.1c}$$

$$Y_{k+1} = Y_{k+\frac{1}{2}} - \frac{\Delta t}{2}\gamma\nabla U(X_{k+1}) + \frac{\Delta t}{2}\lambda^\top Z_{k+1}, \tag{2.3.1d}$$

where Δt denotes the time incremements in the discretisation, ξ_k are i.i.d. standard m-dimensional normal random variables with unit variance and $\theta = 1 - \exp(-\Delta t)$, and $\alpha = \sqrt{1-\theta^2}$. Specifically this is method 2 of [6] applied on a slight modification of (2.1.3), where $\gamma Y_t dt$ and $\gamma\nabla U dt$ is used instead in the r.h.s. of (2.1.3a) and (2.1.3b). Tuning γ can improve numerical perfomance especially in high dimensional problems, but we note that this has no effect in terms of the instantaneous invariant density in (2.2.2); similar to λ and A, γ will not appear in (2.2.2). Unless stated otherwise, in the remainder we will use $\gamma = 1$.

As we will see below the choices for A make a difference in terms of performance. To illustrate this we will use different choices of the form

$$A = \mu A_i;$$

i here is an index for different forms of A. The first choice will be to set $m = n$ and set $A_1 = I_n$ where I_n is $n \times n$ identity matrix. For the rest, we will use $m = 2n$ and pick without rigorous justification

$$A_2 = \begin{pmatrix} 1.9I_n & 0.4I_n \\ 0.1I_n & 0.1I_n \end{pmatrix}, \quad A_3 = \begin{pmatrix} I_n & 0.5I_n \\ 0.5I_n & I_n \end{pmatrix}, \quad (A_4)_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{1}{mn} & \text{otherwise} \end{cases}.$$

Doubling the state space of Z_t relative to X_t, Y_t allows investigating the effect of injecting more noise in the dynamics has to the overall performance and the state space exploration. As per [81] (following [68]), the constraint that the trace of A is uniformly bounded has been used in selecting the above matrices. Note that A_2 does not satisfy the symmetry assumption for the results, but figures for A_2 are displayed in spite of this because there is an interesting improvement in performance for one of the cases below (see Figure 2.3.2 and also others for the sake of comparison). Similarly we will use in each case $\lambda = \bar{\lambda}\lambda_i$

with $\bar{\lambda} > 0$, $\lambda_1 = I_n$ and

$$\lambda_i = \begin{pmatrix} I_n \\ 0 \end{pmatrix}$$

for $i = 2, 3, 4$. As a result $\bar{\lambda}, \mu > 0$ are the main tuning constants for (2.3.1) that do not involve the annealing schedule.

The Langevin system (2.1.2) will be approximated with a leapfrog scheme, that is similar to (2.3.1) in an effort to minimise differences arising from the numerical error,

$$Y_{k+\frac{1}{2}} = Y_k - \hat{\theta}\gamma\nabla U(X_k) - \hat{\theta}\mu Y_k + \hat{\alpha}\sqrt{\mu T_k}\xi_k, \tag{2.3.2a}$$

$$X_{k+1} = X_k + \Delta t\gamma Y_k, \tag{2.3.2b}$$

$$Y_{k+1} = Y_{k+\frac{1}{2}} - \hat{\theta}\gamma\nabla U(X_{k+1}) - \hat{\theta}\mu Y_{k+\frac{1}{2}} + \hat{\alpha}\sqrt{\mu T_{k+1}}\xi_{k+\frac{1}{2}}, \tag{2.3.2c}$$

for $\hat{\theta} = 1 - \exp(-\frac{\Delta t}{2})$, and $\hat{\alpha} = \sqrt{1 - \hat{\theta}^2}$, where in the implementation, (2.3.2a) and (2.3.2c) are combined (aside from the first iteration) and only integer-indexed ξ are used. To make valid comparisons, both (2.3.1) and (2.3.2) will use $\gamma = 1$ and the same noise realisation ξ_k (or the first common n elements) and the same step size Δt.

Finally for both cases we will use following annealing schedule:

$$T_k = \left(\frac{1}{5} + \frac{\ln(1 + k\Delta t)}{E}\right)^{-1},$$

where E is an additional tuning parameter (since $\hat{E}$ is unknown in general).

2.3.2 Sample path properties

Our first set of simulations focus on illustrating some properties of the sample paths generated by (2.3.1) and (2.3.2). We will use the following bivariate potential function as a toy problem

$$U(x_1, x_2) = \frac{x_1^2}{5} + \frac{x_2^2}{10} + 5e^{-x_1^2} - 7e^{-(x_1+5)^2-(x_2-3)^2} - 6e^{-(x_1-5)^2-(x_2+2)^2} + \frac{\frac{2}{3}x_1^2 e^{-\frac{x_1^2}{9}}\cos(x_1 + 2x_2)\cos(2x_1 - x_2)}{1 + \frac{x_2^2}{9}}. \tag{2.3.3}$$

The global minimum is located at $(-5, 3)$, but there are plenty of local minima where the process can get trapped. In addition, there is a barrier along the vertical line $\{x_1 = 0\}$ that makes crossing from each half plane less likely. Here we set $\Delta t = 0.1$, $E = 5$ and each sample is initialised at $(4, 2)$. As a result, it is harder to cross $\{x_1 = 0\}$ to reach the global minimum and it is quite common to get stuck in other local minima such as near

(5, −2). We use the number of crossings on $\{x_1 = 0\}$ as a scale for how stuck the process is in Table 2.3.1. Note that the asymmetric $A = A_2$ case displays the smallest number of crossings.

To illustrate this, in Figure 2.3.1 we present contour plots of U together with a typical realisation of sample paths (in the left panels) for (2.3.2) and (2.3.1) for the different choices of A_i. As expected, (2.3.1) generates smoother paths than those of (2.3.2). We also employ independent runs of each stochastic process for the same initialisation. The results are presented in the right panels of Figure 2.3.1, where we show heat maps for two dimensional histograms representing the frequency of visiting each (x_1, x_2) location over 20 independent realisations of each process. The heat maps in Figure 2.3.1 do not directly depict time dependence in the paths and only illustrate which areas are visited more frequently. Of course converging at the global minumum or the local one at $(5, -2)$ will result in more visits at these areas. The aim here is to investigate the exploration of the state space.

Method equation	Number of transitions across $x = 0$
(2.3.2)	11295
(2.3.1) with $A = A_1$	11893
(2.3.1) with $A = A_2$	10915
(2.3.1) with $A = A_3$	11728
(2.3.1) with $A = A_4$	11771

Tab. 2.3.1: Number of crossings across the vertical line $\{x_1 = 0\}$ for U defined in (2.3.3). The results are summed from $k = 10^5$ iterations of 10^4 independent runs.

2.3.3 Performance and tuning

As expected, the tuning parameters, E, $\bar{\lambda}$ and μ play significant roles in the performance of the discretisations. As E is common to both (2.1.2) and (2.1.3), we wish to demonstrate numerically that the additional tuning variable for the higher order Markovian approximation to the generalised Langevin dynamics improves performance, in spite of the lack of improvement in our theoretical guarantees over Langevin dynamics (2.1.2).

We first comment on relative scaling of $\bar{\lambda}$ and μ based on earlier work for quadratic U and $T_t = T$ being constant. A quadratic U satisfies the bounds in Assumption 1 and is of particular interest because analytical calculations are possible for the spectral gap of L_t, which in turn gives the (exponential) rate of convergence to the equilibrium distribution. It is observed numerically in [152] that in this case, (2.1.3) has a spectral gap that is approximately a function of $\frac{\bar{\lambda}^2}{\mu}$. On the other hand, the spectral gap of (2.1.2) with quadratic U is a function of μ thanks to Theorem 3.1 in [135]. For the rest of the comparison, we will use $\frac{\bar{\lambda}^2}{\mu}$ and μ as variables for the respective discretisations as these quantities appear to have a distinct effect on the mixing in each case. We mention that these choices of variable also allows one to adjust the global Lipschitz constant of the drift coefficient for free in the generalised Langevin equation (2.1.3) up to that of ∇U and 1, whilst in (2.1.2), this grows as μ grows. Therefore one can expect to be able to take a stepsize in the simplest (Euler-Maruyama) discretisation of (2.1.3) that is at least that of (2.1.2) for numerical stability [112, Section 9.8] of the approximation. A detailed stability analysis is beyond the scope of the thesis, we refer the reader to [6] and [112] (see paragraph before Exercise 9.8.1).

We will mainly consider the popular Alpine function in 12 dimensions (see Table 2.3.2, ∇U_1 here is a subgradient), with additional cases presented in Section 2.3.4, setting $\Delta t = 0.02$ (see Section 2.3.1). Note the Alpine function does not strictly satisfy Assumption 1, but since drift conditions for Lyapunov functions are typically available even for weakly growing potentials [52] for the dynamics considered here, the trajectories are expected (and are observed) to remain in a loose sense close to 0. Therefore we may mollify or modify the behaviour at infinity of U to satisfy Assumption 1 with no real observable consequence.

We will initialise at a point well separated from the global minimum and consider each method to be successful if, at the end of the simulation, either the endpoint or an average of the last points are contained within a tolerance region, chosen visually, around the global minumum.

In Figure 2.3.2 we present proportions of 20 independent simulations converging at the region near the global minimum for $U = U_1$ (Table 2.3.2) depending on E and μ for the discretisation of the Langevin dynamics and on E and $\frac{\bar{\lambda}^2}{\mu}$ for that of the generalised Langevin dynamics based on discussion above. Each simulation is run for $k = 5 \cdot 10^4$ iterations. The left panels of Figure 2.3.2 are based on final state and the right on an average of the positions (of X) over the last 5000 iterations. In this example it is clear empirically that the generalised Langevin dynamics result in a higher probability of reaching the global minumum. Another interesting observation is that for the generalised Langevin dynamics good performance is more robust to the chosen value of E. In this example, this means that adding an additional tuning variable and scaling μ proportional to $\bar{\lambda}^2$ makes it easier to find a configuration of the parameters $E, \mu, \bar{\lambda}$ that leads to good perfomance, compared to using the Langevin dynamics and tuning E, μ. It's also worth noting the cases of small E where the generalised Langevin dynamics performs significantly better than the Langevin dynamics in the top plot and even than the case of the same dynamics and larger E. This is an improvement that is not completely encapsulated by the analytic results here; it indicates that the deterministic dynamics ($E = 0$) can be inherently much more successful at climbing out of local minima, which

translates to better convergence rates in the $E > 0$ cases.
The selection $A = A_2$, shown as the middle row in each of Figures 2.3.2, 2.3.4 and 2.3.5, does not satisfy the probably superfluous symmetry assumption as stated in the introduction, but it is noteworthy that the performance varies to such a large extent for different U and that any optimality of A, left as future work, could change depending on whether the symmetry assumption is in place.

2.3.4 Additional cases of U

To produce the figures related to (2.3.1), after setting $E, \frac{\bar{\lambda}^2}{\mu}$ we pick a random value of μ from a grid. The aim of this procedure is to ease visualisation, reduce computational cost and to emphasise that it is $\frac{\bar{\lambda}^2}{\mu}$ that is crucial for mixing and the performance here is not a product of a tedious tuning for μ. U_2 is modified to have the same quadratic confinement in x_1 and x_2 direction and there are several additional local minima due to the last term in the sum. More importantly, compared to (2.3.3) (and U_3) it has a narrow region near the origin that allows easier passage through $\{x_1 = 0\}$. On the other hand U_3 similar to (2.3.3) except that the well near the global minimum (and the dominant local minimum at $(5, -2)$) are elongated in the direction of x_2 (and x_1 respectively).

Cost function	Initial condition	Tolerance sets
$U_1(x) = \frac{1}{2}\sum_{i=1}^{12} \lvert x_i \sin(x_i) + 0.1x_i \rvert$	$x_j = 6\ \forall j$	$x_j \in [-2, 2]\ \forall j$
$U_2(x_1, x_2) = \frac{x_1^2}{7} + \frac{x_2^2}{7} + 5\left(1 - e^{-9x_2^2}\right)e^{-x_1^2}$ $-7e^{-(x_1+5)^2-(x_2-3)^2}$ $-6e^{-(x_1-5)^2-(x_2+2)^2}$ $+\frac{\frac{2}{3}x_1^2 e^{-\frac{x_1^2}{9}} \cos(x_1+2x_2)\cos(2x_1-x_2)}{1+\frac{x_2^2}{9}}$	$x_1 = 4$, $x_2 = 2$	$x_1 \in [-6.5, -4.5]$, $x_2 \in [1.5, 4.5]$
$U_3(x_1, x_2) = \frac{x_1^2}{5} + \frac{x_2^2}{10} + 5e^{-x_1^2}$ $-7e^{-2(x_1+5)^2-\frac{(x_2-3)^2}{5}}$ $-6e^{-\frac{(x_1-5)^2}{5}-2(x_2+2)^2}$	$x_1 = 4$, $x_2 = 2$	$x_1 \in [-6.5, -4.5]$, $x_2 \in [1.5, 4.5]$

Tab. 2.3.2: Details of three different cost functions, initialisation and tolerance regions corresponding to regions of attraction of the global minimum.

In Figures 2.3.4 and 2.3.5 we present results for U_2 and U_3. A notable difference to Figure 2.3.2 here is that the panels on the left show proportions of the position average of the last 5000 iterations being near the correct global minimum and the panels on the right present the number of jumps across $\{x_1 = 0\}$ demonstrated by a position average at each iteration of the previous 5000 iterations. More precisely, the panels on the right show the number of jumps shown by the trajectory

$$\tilde{X}_k = \frac{1}{5000} \sum_{k'=1}^{5000} X_{k-k'+1}$$

for all $k > 5000$. All results are averaged over 20 independent runs. The aim here is to measure the extent of exploration of each process similar to Table 2.3.1. We observe that in both cases using (2.3.1) leads to a similar number of jumps. We believe the benefit of the higher order dynamics here are the robustness of performance for different values of E and $\frac{\bar{\lambda}^2}{\mu}$. This is especially for using A_3 and A_4. Finally we note that despite similarities between U_2 and U_3 there are significant features that are different: the sharpness in the confinement, the shape and number of attracting wells and the shape of barriers that obstruct crossing regions in the state space. This will have a direct effect in performance, which can explain the difference in performance when comparing Figures 2.3.4 and 2.3.5; U_3 is a harder cost function to minimise.
We complement our results with an application of (2.3.1) on the optimisation of parameters in a neural network with respect to a loss function in Figure 2.3.3. In particular, (2.3.1) with $\bar{\lambda} = \mu = 1$, $E = 0.001$ and $\Delta t = 0.05$ is used to update the parameters in a neural network of size $n = 122970$. Here, the function U is fixed as the cross entropy between the network prediction and target values for the MATLAB Digits data set[2] and full gradient evaluations are used. We follow the network architecture and overall implementation found in https://uk.mathworks.com/help/deeplearning/ref/dlupdate.html. Although a thorough comparison between different parameter choices and dynamics for such a high dimensional problem is beyond the scope of the thesis, Figure 2.3.3 illustrates the feasibility of (2.3.1) on such problems.

2.4 Proofs

2.4.1 Notation and preliminaries

In this section, unless stated otherwise, ∂_t is used to denote the partial derivative with respect to t with T_t fixed (whenever its operand depends on T_t), whereas $\frac{d}{dt}$ denotes the full derivative in t. In addition, ∇ denotes the gradient in $\mathbb{R}^{2n+m}$ space and $d\zeta$ will be used for the Lebesgue measure on $\mathbb{R}^{2n+m}$. The notation $\mathbb{1}_S$ will be used for the indicator function on the set S.

[2] https://uk.mathworks.com/help/deeplearning/ug/data-sets-for-deep-learning.html

For all $k > 0$, recall the standard mollifier $\varphi : \mathbb{R} \to \mathbb{R}$ and $\varphi_k : \mathbb{R} \to \mathbb{R}$ to be:

$$\varphi_k(x) := \frac{1}{k}\varphi\bigg(\frac{x}{k}\bigg), \qquad \varphi(x) := \begin{cases} e^{\frac{1}{x^2-1}}\bigg(\int_{-1}^{1} e^{\frac{1}{y^2-1}}\,dy\bigg)^{-1} & \text{if } -1 < x \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.4.1}$$

For existence and uniqueness of (2.1.3), we will use the setting in [163]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $\mathcal{F}_t$, $t \in [0, \infty)$ be a normal filtration. Here $(W_t)_{t\ge 0}$ is a standard Wiener process on $\mathbb{R}^m$ with respect to $\mathcal{F}_t$, $t \in [0, \infty)$.
The formal[3] $L^2(\mu_{T_t})$-adjoint L_t^* of L_t is given by

$$L_t^* = -\left(y \cdot \nabla_x - \nabla_x U(x) \cdot \nabla_y\right) - \left(z^\top \lambda \nabla_y - y^\top \lambda^\top \nabla_z\right) - T_t^{-1} z^\top A \nabla_z + A : D_z^2. \tag{2.4.2}$$

Let $\epsilon \ge 0$ and consider the perturbed system

$$dX_t^\epsilon = Y_t^\epsilon\,dt + \epsilon(-T_t^{-1}\nabla_x U(X_t^\epsilon)\,dt + dW_t^1), \tag{2.4.3a}$$
$$dY_t^\epsilon = -\nabla_x U(X_t^\epsilon)\,dt + \lambda^\top Z_t^\epsilon\,dt + \epsilon(-T_t^{-1} Y_t^\epsilon\,dt + dW_t^2), \tag{2.4.3b}$$
$$dZ_t^\epsilon = -\lambda Y_t^\epsilon\,dt - T_t^{-1} A Z_t^\epsilon\,dt + \Sigma\,dW_t^3, \tag{2.4.3c}$$

with $(X_0^\epsilon, Y_0^\epsilon, Z_0^\epsilon) = (X_0, Y_0, Z_0)$ restricted as in Assumption 4, where W_t^1, W_t^2, W_t^3 are independent n-dimensional and m-dimensional Wiener processes. As before, the law and density of (2.4.3) will be denoted by m_t^ϵ along with $h_t^\epsilon = \frac{dm_t^\epsilon}{d\mu_{T_t}}$. Let the linear differential operators S_t^x, S_t^y and their respective formal L^2-adjoints $S_t^{x\top}$ and $S_t^{y\top}$ be given by

$$S_t^x = -T_t^{-1}\nabla_x U \cdot \nabla_x + \Delta_x, \qquad S_t^y = -T_t^{-1} y \cdot \nabla_y + \Delta_y,$$
$$S_t^{x\top} = T_t^{-1}\nabla_x U \cdot \nabla_x + T_t^{-1}\Delta_x U + \Delta_x, \qquad S_t^{y\top} = T_t^{-1} y \cdot \nabla_y + T_t^{-1} n + \Delta_y.$$

Note that the formal $L^2(\mu_{T_t})$-adjoints of S_t^x and S_t^y coincide with S_t^x and S_t^y, so that the generator, denoted L_t^ϵ, associated to (2.4.3) and its formal $L^2(\mu_{T_t})$-adjoint are given by the formal operators

$$L_t^\epsilon = L_t + \epsilon(S_t^x + S_t^y), \qquad L_t^{\epsilon *} = L_t^* + \epsilon(S_t^x + S_t^y).$$

For any $\phi \in \mathcal{C}^\infty$ and $f : \mathbb{R}^{2n+m} \to \mathbb{R}$ smooth enough,

$$L_t^\epsilon(\phi(f)) = \phi'(f) L_t^\epsilon(f) + \phi''(f)\Gamma_t^\epsilon(f), \tag{2.4.4}$$

[3] See for instance Appendix B in [62]. In the present chapter the infinitesimal generators and their adjoints are considered as honest differential operators acting on smooth functions.

where Γ_t^ϵ is the carré du champ operator for L_t^ϵ given by

$$\Gamma_t^\epsilon(f) = \frac{1}{2}L_t^\epsilon(f^2) - fL_t^\epsilon(f) = \nabla f \cdot (A^\epsilon \nabla f), \tag{2.4.5}$$

$A^\epsilon \in \mathbb{R}^{(2n+m)\times(2n+m)}$ denotes the matrix with entries

$$A_{ij}^\epsilon := \begin{cases} \epsilon & \text{if } 1 \le i = j \le 2n, \\ A_{i-2n,j-2n} & \text{if } 2n+1 \le i,j \le 2n+m, \\ 0 & \text{otherwise} \end{cases}$$

and $A_{i,j}$ denotes the $(i,j)^{\text{th}}$ entry of A. Let $\mathcal{C}_+^\infty = \{f \in \mathcal{C}^\infty : f > 0\}$. For $\Phi : \mathcal{C}_+^\infty \to \mathcal{C}^\infty$ differentiable in the sense that for any $f \in \mathcal{C}_+^\infty$, $g \in \mathcal{C}^\infty$,

$$(d\Phi(f).g)(\zeta) \quad := \quad \lim_{s\to 0} \frac{(\Phi(f+sg))(\zeta) - (\Phi(f))(\zeta)}{s}$$

exists for all $\zeta \in \mathbb{R}^{2n+m}$, the Γ_Φ operator for $L_t^{\epsilon*}$ is defined by

$$\Gamma_{L_t^{\epsilon*},\Phi}(h) \quad := \quad \frac{1}{2}(L_t^{\epsilon*}\Phi(h) - d\Phi(h).(L_t^{\epsilon*}h)). \tag{2.4.6}$$

As is well-known, $L_t^{\epsilon*}$ does not satisfy the standard chain and product rules due to the additional term from the second derivatives in $L_t^{\epsilon*}$; straightforward calculations give:

$$L_t^{\epsilon*}(\psi(f)) \;=\; \psi'(f)L_t^{\epsilon*}f + \psi''(f)\nabla f \cdot (A^\epsilon \nabla f) \tag{2.4.7}$$

$$L_t^{\epsilon*}(fg) \;=\; fL_t^{\epsilon*}(g) + gL_t^{\epsilon*}(f) + \nabla f \cdot (2A^\epsilon \nabla g) \tag{2.4.8}$$

for all $f, g \in \mathcal{C}^\infty$ and $\psi \in \mathcal{C}^\infty$. Note $\nabla f \cdot (A^\epsilon \nabla f)$ and $\nabla f \cdot (2A^\epsilon \nabla g)$ are respectively the *carré du champ* and its symmetric bilinear operator via polarisation for $L_t^{\epsilon*}$.
In addition, for a scalar-valued D_1 and a vector-valued operator D_2 both acting on scalar-valued functions, denote the commutator bracket as follows:

$$[D_1, D_2]h = (D_1(D_2h)_1 - (D_2D_1h)_1,\ \ldots, D_1(D_2h)_{d_{D_2}} - (D_2D_1h)_{d_{D_2}}) \tag{2.4.9}$$

for $h \in \mathcal{C}^\infty$, where $d_{D_2} \in \mathbb{N}$ is the number of elements in the output of D_2.

2.4.2 Auxiliary results

For the next result, the space of smooth functions that will be used is from [38]: let $\mathcal{C}_{b,c}^\infty = \mathcal{C}_{b,c}^\infty((0,\infty) \times \mathbb{R}^{2n+m})$ be the space of real-valued functions $f : (0,\infty) \times \mathbb{R}^{2n+m} \to \mathbb{R}$ such that

1. f is measurable with respect to $\mathcal{B}((0,\infty)) \otimes \mathcal{B}(\mathbb{R}^{2n+m})$,

2. for all $t > 0$, $f(t,\cdot)$ is smooth and f is bounded on compact subsets of $\mathbb{R}_{>0} \times \mathbb{R}^{2n+m}$.

Proposition 2.4.1. *Under Assumption 1, 3 and 4, for all $t > 0$ and $\epsilon \geq 0$, the unique strong solution $(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon)$ to (2.4.3) is well-defined and there exists some constant $\kappa > 0$ such that*

$$\mathbb{E}\big[|X_t^\epsilon|^2 + |Y_t^\epsilon|^2 + |Z_t^\epsilon|^2\big] \leq e^{\kappa t}\mathbb{E}\big[|X_0|^2 + |Y_0|^2 + |Z_0|^2\big] < \infty. \tag{2.4.10}$$

Furthermore, for all time $t > 0$, the law of the process $(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon)$

- *admits an almost-everywhere finite strictly positive density, also denoted m_t^ϵ, with respect to the Lebesgue measure on $\mathbb{R}^{2n+m}$,*

- *is the unique integrable distributional solution to the Fokker-Planck-Kolmogorov equation [20]*

$$\begin{cases} \partial_t m_t^\epsilon = (L_t^\top + \epsilon(S_t^{x\top} + S_t^{y\top}))m_t^\epsilon \\ m_0^\epsilon = m_0, \end{cases} \tag{2.4.11}$$

  where $L_t^\top$ is the formal L^2-adjoint of L_t.

Finally when $\epsilon > 0$, $m_\bullet$ and its partial derivative in time belongs in $\mathcal{C}_{b,c}^\infty$.

For the notion of integrable distributional solutions, see p.338 in [20].

Proof. Existence and uniqueness of an almost surely continuous $\mathcal{F}_t$-adapted processes follows by conditions (2.2.3) and (2.2.5) using Theorem 3.1.1 in [163]; in addition, (2.4.10) holds by the same theorem. For the claim that the law admits a density, we will apply Theorem 1 in [94] for the case of an arbitrary deterministic starting point. First, condition (H1) in the same article is verified. Take the sets 'K_n' to be

$$K_p = \prod_{i=1}^{2n+m} [-p, p]$$

for all $p \in \mathbb{N}$. The unique solution to (2.4.3) with a deterministic starting point

$$(X_0, Y_0, Z_0) = (x_0, y_0, z_0) \in \mathbb{R}^{2n+m}$$

satisfies the same bound (2.4.10) as before when initialising from m_0. Moreover, for the random sets

$$\Xi_p = \{s > 0 : (X_u^\epsilon, X_u^\epsilon, X_u^\epsilon) \in K_p, 0 \leq u \leq s\},$$

for $p \in \mathbb{N}$, the solution $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \hat{Z}_t^\epsilon)$ to the stopped stochastic differential equation

$$
\begin{aligned}
d\hat{X}_t^{\epsilon,p} &= \mathbb{1}_{\Xi_p}(t)(\hat{Y}_t^{\epsilon,p}\,dt + \epsilon(-T_t^{-1}\nabla_x U(\hat{X}_t^{\epsilon,p})\,dt + dW_t^1)), && \text{(2.4.12a)}\\
d\hat{Y}_t^{\epsilon,p} &= \mathbb{1}_{\Xi_p}(t)(-\nabla_x U(\hat{X}_t^{\epsilon,p})\,dt + \lambda^\top \hat{Z}_t^{\epsilon,p}\,dt + \epsilon(-T_t^{-1}\hat{Y}_t^{\epsilon,p}\,dt + dW_t^2)), && \text{(2.4.12b)}\\
d\hat{Z}_t^{\epsilon,p} &= \mathbb{1}_{\Xi_p}(t)(-\lambda\hat{Y}_t^{\epsilon,p}\,dt - T_t^{-1}A\hat{Z}_t^{\epsilon,p}\,dt + \Sigma\,dW_t^3), && \text{(2.4.12c)}
\end{aligned}
$$

is well-defined by the same Theorem 3.1.1 in [163] and the corresponding bound

$$
\mathbb{E}\big[|\hat{X}_t^{\epsilon,p}|^2 + |\hat{Y}_t^{\epsilon,p}|^2 + |\hat{Z}_t^{\epsilon,p}|^2\big] \le e^{\kappa t}\big(|x_0|^2 + |y_0|^2 + |z_0|^2\big) < \infty
$$

holds. Identifying $(\hat{X}_t^{\epsilon,p}, \hat{Y}_t^{\epsilon,p}, \hat{Z}_t^{\epsilon,p}) = (X^\epsilon_{t\wedge \sup \Xi_p}, Y^\epsilon_{t\wedge \sup \Xi_p}, Z^\epsilon_{t\wedge \sup \Xi_p})$ a.s. yields that[4] for any $\tau > 0$

$$
\begin{aligned}
\mathbb{P}(\inf\{t \ge 0 : (X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon) \notin K_p\} \le \tau) &\le \frac{1}{p^2}\mathbb{E}\big[|X^\epsilon_{\tau\wedge\sup\Xi_p}|^2 + |Y^\epsilon_{\tau\wedge\sup\Xi_p}|^2 + |Z^\epsilon_{\tau\wedge\sup\Xi_p}|^2\big]\\
&\le \frac{e^{\kappa\tau}}{p^2}\big(|x_0|^2 + |y_0|^2 + |z_0|^2\big)
\end{aligned}
$$

and in particular that for any $\tau > 0$,

$$
\mathbb{P}(\inf\{t \ge 0 : (X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon) \notin K_p\} \le \tau) \to 0 \quad \text{as } p \to \infty. \tag{2.4.13}
$$

Suppose for contradiction that with nonzero probability, the increasing-in-p random variable $\inf\{t \ge 0 : (X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon) \notin K_p\}$ converges to a real value as $p \to \infty$. Then there exists a time $\hat{\tau} > 0$ such that with nonzero probability,

$$
\inf\{t \ge 0 : (X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon) \notin K_p\} \le \hat{\tau} \quad \forall p \in \mathbb{N},
$$

which contradicts (2.4.13). Therefore condition (H1) in [94] holds for (2.4.3). Condition (H2) in the same article holds due the K_p being compact and the smoothness assumption on U. It can be readily checked that the local weak Hörmander condition, see definition in [94, (LWH)], also holds at any (t, y_0) for any $r \in (0, t)$ and $R > 0$. Therefore[5] by Theorem 1 in [94], due to our Assumptions 1 and 3, the solution to (2.4.3) with a deterministic starting point $\zeta_0 \in \mathbb{R}^{2n+m}$ admits a smooth density $p_t^{\zeta_0} \in \mathcal{C}^\infty(\mathbb{R}^{2n+m})$ for all $t > 0$. Moreover by Theorem 2 in [94], for any fixed $\zeta \in \mathbb{R}^{2n+m}$, $\mathbb{R}^{2n+m} \ni \zeta_0 \mapsto p_t^{\zeta_0}(\zeta)$ is lower semi continuous and hence measurable, so that the $\mathbb{R} \cup \{\pm\infty\}$-valued function

[4] Alternatively Corollary 1.2 of Section 5 in [67] can be used.

[5] Note that there is a wealth of related results, e.g. [30, 46], but [94] seems to contain the only immediately applicable (and relevant) result for our particular case; see also remark 2.4.1.

on $\mathbb{R}^{2n+m}$,

$$\int_{\mathbb{R}^{2n+m}} p_t^{\zeta_0} m_0(d\zeta_0), \tag{2.4.14}$$

is integrable by Fubini's theorem and so is almost everywhere $\mathbb{R}$-valued on $\mathbb{R}^{2n+m}$. By Itô's rule, (2.4.14) solves (2.4.11) in the distributional sense. In addition, (2.4.11) is the unique integrable solution by Theorem 9.6.3 in [20], which requires for any $T > 0$ that there exists $V \in C^2(\mathbb{R}^{2n+m})$ such that

1. $V(x) \to \infty$ as $|x| \to \infty$ and
2. for some constant $C_V > 0$ and all $(x,t) \in \mathbb{R}^{2n+m} \times (0,T)$, it holds that $L_t^\epsilon V \geq -C_V V$ and $|\nabla V| \leq C_V V$.

Setting $V(x,y,z) = 1 + U(x) - U_m + \frac{|y|^2}{2} + \frac{|z|^2}{2}$ and calculating

$$L_t^\epsilon\left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2}\right) = \epsilon\left(-\frac{1}{T_t}|\nabla_x U|^2 + \Delta_x U - \frac{1}{T_t}|y|^2 + n\right) - \frac{1}{T_t} z^\top A z + \mathrm{Tr} A, \tag{2.4.15}$$

it is clear from assumptions (2.2.3), (2.2.5) and either (2.2.6) or (2.2.7) on U that these conditions are satisfied since T is finite; therefore there is a unique integrable solution to (2.4.11) in the sense of p.338 in [20]. The expression in (2.4.14) is thus the density for the law of the solution to (2.4.3) with initial law m_0 at time t.
For $\epsilon > 0$, the time-depending law of $(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon)$ and its partial derivative with respect to time belongs in $\mathcal{C}_{b,c}^\infty$ by Theorem 1.1 in [38] because (2.4.14) satisfies (2.4.11).
For positivity of the density where $\epsilon = 0$, the steps in Lemma 3.4 of [132] involving the solution to an associated control problem can be followed. The associated control problem has the expression

$$\frac{d}{dt}\begin{pmatrix} Q_t \\ P_t \\ V_t \end{pmatrix} = \begin{pmatrix} P_t \\ -\nabla U(Q_t) + \lambda^\top V_t \\ -\lambda P_t - T_t A V_t + \Sigma \frac{d\tilde{U}}{dt} \end{pmatrix}. \tag{2.4.16}$$

It suffices to show that given any $S > 0$ and any pair $(Q_0, P_0, V_0) \in \mathbb{R}^{2n+m}$ and $(Q^*, P^*, V^*) \in \mathbb{R}^{2n+m}$, there exists a control $\tilde{U} : [0,\infty) \to \mathbb{R}^m$ such that the solution (Q_t, P_t, V_t) to (2.4.16) starting at (Q_0, P_0, V_0) satisfies $(Q_S, P_S, V_S) = (Q^*, P^*, V^*)$. Fix $S > 0, \epsilon > 0, (Q_0, P_0, V_0) \in \mathbb{R}^{2n+m}$, $(Q^*, P^*, V^*) \in \mathbb{R}^{2n+m}$. Using the mollifier (2.4.1), let $\nu := \varphi_{\frac{1}{2}} * \mathbb{1}_{(-\infty, \frac{1}{2}]}$, where $*$ denotes convolution. Define a smooth function $\hat{Q}_\cdot : [0,S] \to \mathbb{R}^n$

by

$$\hat{Q}_t = \left((-\nabla U(Q_0) + \lambda^\top V_0)\frac{t^2}{2} + P_0 t + Q_0\right)\nu\left(\frac{t}{S}\right) \\ + \left((-\nabla U(Q^*) + \lambda^\top V^*)\frac{t^2}{2} + P^* t + Q^*\right)\nu\left(1 - \frac{t}{S}\right), \tag{2.4.17}$$

which satisfies

$$\hat{Q}_0 = Q_0, \qquad \hat{Q}_S = Q^*, \qquad \frac{d\hat{Q}_t}{dt}(0) = P_0, \qquad \frac{d\hat{Q}_t}{dt}(S) = P^*.$$

Define $\hat{P}_\cdot : [0, S] \to \mathbb{R}^n$ through

$$\hat{P}_t = \frac{d\hat{Q}_t}{dt}. \tag{2.4.18}$$

For $\hat{V}_\cdot : [0, S] \to \mathbb{R}^m$, $\hat{V}_\cdot : [0, S] \to \mathbb{R}^m$ is defined with

$$\hat{V}_t = \lambda(\lambda^\top \lambda)^{-1}\Bigg(\nabla U(\hat{Q}_t) + \partial_t^2\Bigg[\left(-\nabla U(Q_0)\frac{t^2}{2} + P_0 t + Q_0\right)\nu\left(\frac{t}{S}\right) \\ + \left(-\nabla U(Q^*)\frac{t^2}{2} + P^* t + Q^*\right)\nu\left(1 - \frac{t}{S}\right)\Bigg]\Bigg) + \partial_t^2\left[V_0\frac{t^2}{2}\nu\left(\frac{t}{S}\right) + V^*\frac{t^2}{2}\nu\left(1 - \frac{t}{S}\right)\right], \tag{2.4.19}$$

where $(\lambda^\top \lambda)^{-1}$ exists by λ having rank n. Note that $\hat{V}_t$ satisfies $\hat{V}_0 = V_0$ and $\hat{V}_S = V^*$. Let the smooth function $\tilde{U} : [0, \infty) \to \mathbb{R}^m$ be given by

$$\frac{d\tilde{U}}{dt} = \Sigma^{-1}\left(\frac{d\hat{V}_t}{dt} + \lambda\hat{P}_t + T_t A\hat{V}_t\right), \quad \tilde{U}(0) = 0. \tag{2.4.20}$$

For this $\tilde{U}$, the solution to (2.4.16) with initial condition (Q_0, P_0, V_0) is $(\hat{Q}_t, \hat{P}_t, \hat{V}_t)$ by construction; its uniqueness is guaranteed by considering the system satisfied by the difference between two supposedly different solutions (Q_t^1, P_t^1, V_t^1) and (Q_t^2, P_t^2, V_t^2)

$$\frac{d}{dt}\begin{pmatrix} Q_t^1 - Q_t^2 \\ P_t^1 - P_t^2 \\ V_t^1 - V_t^2 \end{pmatrix} = \begin{pmatrix} P_t^1 - P_t^2 \\ -\nabla U(Q_t^1) - \nabla U(Q_t^2) + \lambda^\top(V_t^1 - V_t^2) \\ -\lambda(P_t^1 - P_t^2) - T_t A(V_t^1 - V_t^2) \end{pmatrix}$$

and the time derivative of $|Q_t^1 - Q_t^2|^2 + |P_t^1 - P_t^2|^2 + |V_t^1 - V_t^2|^2$, using (2.2.3) and the mean value theorem on $|\nabla U(Q_t^1) - \nabla U(Q_t^2)|^2$.

With non-zero probability, the path of Brownian motion stays within an ϵ-neighbourhood of any continuously differentiable path, in particular of $\tilde{U}$. Positivity of m_t follows by the support theorem of Stroock and Varadhan (Theorem 5.2 in [182]). The above construction for the $\epsilon > 0$ case follows with a simple modification; equation (2.4.16) becomes

$$\frac{d}{dt}\begin{pmatrix} Q_t \\ P_t \\ V_t \end{pmatrix} = \begin{pmatrix} P_t - \epsilon\nabla U(Q_t) + \epsilon\frac{d\tilde{U}_1}{dt} \\ -\nabla U(Q_t) + \lambda^\top V_t - \epsilon P_t + \epsilon\frac{d\tilde{U}_2}{dt} \\ -\lambda P_t - T_t A V_t + \Sigma\frac{d\tilde{U}}{dt} \end{pmatrix}, \tag{2.4.21}$$

so setting $\frac{d\tilde{U}_1}{dt} = \nabla U(\hat{Q}_t)$ and $\frac{d\tilde{U}_2}{dt} = \hat{P}_t$ together with (2.4.20) gives that (2.4.17), (2.4.18) and (2.4.19) solves equation (2.4.21) and concludes the proof.

□

Remark 2.4.1. For smoothness of the density, the results in [186] can also be considered, but there the assumptions are slightly mismatched. Firstly, the statement assumes boundedness of $\partial^\alpha V$ for any multiindex α where V would in the case here be any of the coefficients appearing in (2.1.3), which fails for $|\alpha| = 0$. Secondly, in case of (A.1) (in [186]), condition (i) fails and in case of (A.2), condition (i) fails due to V_0. Both of these assumptions seem possibly unnecessary in the proofs but we avoid this in favour of the more recent work [94].

The results below up to Proposition 2.4.10 are directed towards showing dissipation of a distorted entropy as required in the proof of Theorem 2.2.5.

2.4.3 Lyapunov function

Lemma 2.4.2. *Under Assumption 1, 3 and 4, there exist constants* $a, b, c, d, \delta > 0$ *independent of* ϵ *such that* $R : \mathbb{R}^{2n+m+1} \to \mathbb{R}$ *defined as*

$$R(x, y, z, T_t) \quad := \quad U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} + \delta T_t\Big(y^\top\lambda^{-1}z + \frac{1}{2}x\cdot y\Big) \tag{2.4.22}$$

satisfies

$$a(|x|^2 + |y|^2 + |z|^2) - d \quad \le \quad R(x, y, z, T_t) \quad \le \quad b(|x|^2 + |y|^2 + |z|^2) + d \tag{2.4.23}$$

and there exists $0 < \epsilon' \le 1$ *for which* $\epsilon \le \epsilon'$ *implies*

$$L_t^\epsilon R \le -cT_t R + \frac{d}{T_t}. \tag{2.4.24}$$

Proof. By the quadratic assumption (2.2.7) on U and boundedness Assumption 3 on T_t, it is clear that there exists $\hat{\delta} > 0$ such that the first statement (2.4.23) holds with $d = \max(|U_m|, |U_M|)$ for all $\delta \in (0, \hat{\delta}]$. Inequality (2.4.24) follows by the following calculation using our assumptions on U, T_t and applications of Young's inequality.

Fix $\delta > 0$ to be

$$\delta \leq \min\Bigg(\hat{\delta}, 1, \frac{4r_1^2}{(r_2+1)\sup_{s\geq 0} T_s}, 2\Big(\sup_{s\geq 0} T_s\Big)^{-1},$$
$$\frac{A_c}{2}\Bigg[\bigg(\frac{|\lambda|^2}{2r_1} + 1 + \frac{r_2}{r_1}\big|\lambda^{-1}\big|^2\bigg)\Big(\sup_{s\geq 0} T_s\Big)^2 + 2(|A|^2+1)\big|\lambda^{-1}\big|^2\Bigg]^{-1}\Bigg), \tag{2.4.25}$$

where $|\cdot|$ is the operator norm here and $A_c > 0$ is the coercivity constant of the positive definite matrix A. Consider each of the terms of $L_t^\epsilon(R)$ seperately.

$$L_t^\epsilon\bigg(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2}\bigg) = \epsilon\bigg(-\frac{1}{T_t}|\nabla_x U|^2 + \Delta_x U - \frac{1}{T_t}|y|^2 + n\bigg) - \frac{1}{T_t} z^\top A z + \mathrm{Tr}A \tag{2.4.26}$$
$$\leq \epsilon\bigg(-\frac{1}{T_t}(r_1^2|x|^2 - 2r_1 U_g) + n|D_x^2 U|_\infty - \frac{1}{T_t}|y|^2 + n\bigg) - \frac{1}{T_t} z^\top A z + \mathrm{Tr}A, \tag{2.4.27}$$

where the last inequality follows from (2.2.4) and $\nabla_x U \cdot x \leq \frac{1}{2r_1}|\nabla_x U|^2 + \frac{r_1}{2}|x|^2$. Using the quadratic bound (2.2.5) on $\nabla_x U$, we get

$$L_t^\epsilon(y^\top \lambda^{-1} z) = -\epsilon T_t^{-1} y^\top \lambda^{-1} z - \nabla_x U \lambda^{-1} z + |z|^2 - |y|^2 - T_t^{-1} z^\top A (\lambda^{-1})^\top y \tag{2.4.28}$$
$$\leq \frac{|y|^2}{4} + 2T_t^{-2}(|A|^2 + \epsilon^2)\big|\lambda^{-1}\big|^2|z|^2 + \frac{r_1}{4r_2}|\nabla_x U|^2 + \frac{r_2}{r_1}\big|\lambda^{-1}\big|^2|z|^2 + |z|^2 - |y|^2$$
$$\leq \frac{r_1}{4}|x|^2 + \frac{r_1}{4r_2}U_g - \frac{3}{4}|y|^2 + \bigg(1 + \frac{r_2}{r_1}\big|\lambda^{-1}\big|^2 + 2T_t^{-2}(|A|^2+1)\big|\lambda^{-1}\big|^2\bigg)|z|^2. \tag{2.4.29}$$

Then using also (2.2.4) for $\nabla_x U \cdot x$, we get

$$L_t^\epsilon(x\cdot y) = -\epsilon T_t^{-1} y \cdot \nabla_x U - \epsilon T_t^{-1} x \cdot y + |y|^2 - \nabla_x U \cdot x + z^\top \lambda x \tag{2.4.30}$$
$$\leq \epsilon T_t^{-1}\bigg(\Big(\frac{r_2}{2} + \frac{1}{2}\Big)|x|^2 + |y|^2 + \frac{U_g}{2}\bigg) + |y|^2 - r_1|x|^2 + U_g + \frac{\big|\lambda^\top\big|^2}{r_1}|z|^2 + \frac{r_1}{4}|x|^2. \tag{2.4.31}$$

Combining (2.4.22), (2.4.27), (2.4.29), (2.4.31) and taking $\epsilon \leq 1$,

$$
\begin{aligned}
L_t^\epsilon(R(x,y,z,T_t)) &= L_t^\epsilon\left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2}\right) + \delta T_t L_t(y^\top \lambda^{-1} z) + \frac{\delta T_t}{2} L_t(x \cdot y) \quad (2.4.32)\\
&\leq -\delta T_t \frac{r_1}{8}|x|^2 - \delta T_t \frac{1}{4}|y|^2 - \frac{1}{T_t} z^\top A z + C + \frac{2\epsilon r_1 U_g}{T_t}\\
&\quad + \delta T_t \left[\frac{\left|\lambda^\top\right|^2}{2r_1} + \left(1 + \frac{r_2}{r_1}\left|\lambda^{-1}\right|^2 + 2T_t^{-2}(|A|^2+1)\left|\lambda^{-1}\right|^2\right)\right]|z|^2.\\
&\quad + \epsilon T_t^{-1}\left[\left(\frac{\delta T_t}{4}(r_2+1) - r_1^2\right)|x|^2 + \left(\frac{\delta T_t}{2} - 1\right)|y|^2\right], \quad (2.4.33)
\end{aligned}
$$

where $0 < C = n(|D_x^2 U|_\infty + 1) + \mathrm{Tr}A + \delta \frac{r_1 U_g}{4r_2} \sup_{s\geq 0} T_s + \delta U_g(\frac{1}{4} + \frac{\sup_{s\geq 0} T_s}{2})$. Therefore for δ satisfying the bound (2.4.25), the first square bracket term satisfies

$$
\delta T_t \left[\frac{\left|\lambda^\top\right|^2}{2r_1} + \left(1 + \frac{r_2}{r_1}\left|\lambda^{-1}\right|^2 + 2T_t^{-2}(|A|^2+1)\left|\lambda^{-1}\right|^2\right)\right] \leq \frac{1}{2}\left(\sup_{s\geq 0} T_s\right)^{-1} A_c |z|^2
$$

and the second square bracket term is negative, where the assumption that T_t is bounded above for all time has been used. Rearranging,

$$
\begin{aligned}
L_t^\epsilon(R(x,y,z,T_t)) &\leq -\delta T_t \frac{r_1}{8}|x|^2 - \delta T_t \frac{1}{4}|y|^2 - \frac{A_c}{2\sup_s T_s}|z|^2 + C + \frac{2\epsilon r_1 U_g}{T_t}\\
&\leq -cT_t R + C' T_t^{-1},
\end{aligned}
$$

where $c > 0$ is small enough, $C' > 0$ is large enough and the right inequality of (2.4.23) has been used. The result follows using $d = \max(C', |U_m|, |U_M|)$.

□

Lemma 2.4.3. *Under Assumption 1, 3, 4 and for $0 \leq \epsilon \leq \epsilon'$, the solution $(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon)$ to (2.4.3) is such that $\frac{\mathbb{E}[R(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon, T_t)]}{(\ln(e+t))^2}$ is bounded uniformly in time t and in ϵ.*

Proof. It is equivalent to prove the result for $R + d > 0$ in place of R. Let $R_t := R(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon, T_t)$. Firstly, by (2.4.22), the left hand bound in (2.4.23) and the Assumption 3,

$$
T_t' \mathbb{E}\left[\delta\left((Y_t^\epsilon)^\top \lambda^{-1} Z_t + \frac{1}{2} X_t^\epsilon \cdot Y_t^\epsilon\right)\right] \leq |T_t'| \mathbb{E}\left[\delta\left|(Y_t^\epsilon)^\top \lambda^{-1} Z_t^\epsilon + \frac{1}{2} X_t^\epsilon \cdot Y_t^\epsilon\right|\right] \leq \frac{B}{t}\mathbb{E}[R_t + d] \tag{2.4.34}
$$

for a constant $B \geq 0$ independent of ϵ. By Itô's rule and for $t_0 < s < t$,

$$\mathbb{E}R_t - \mathbb{E}R_s = \int_s^t \mathbb{E}(T'_u \partial_{T_t} R + L_u^\epsilon R)(X_u^\epsilon, Y_u^\epsilon, Z_u^\epsilon, T_u) du,$$

where (2.4.34), (2.4.32) (2.4.26), (2.4.28), (2.4.30), (2.2.3), (2.2.5) have been used together with Fubini's theorem. Property (2.4.24) from Lemma 2.4.2 and (2.4.34) give

$$\begin{aligned} \mathbb{E}[R_t + d] - \mathbb{E}[R_s + d] &\leq \int_s^t \left(\frac{B}{u} \mathbb{E}[R_u + d] + \mathbb{E}[-cT_u R_u + dT_u^{-1}] \right) du \\ &\leq \int_s^t \left(\left(\frac{B}{u} - cT_u \right) \mathbb{E}[R_u + d] + B' T_u^{-1} \right) du \end{aligned} \tag{2.4.35}$$

for a constant $B' \geq 0$ independent of ϵ. By the spatial quadratic bounds on R and (2.4.10), the right-hand side is converging to zero as $s \to t$. Since a similar lower bound of the left-hand side may be obtained by explicit calculations as before for the upper bound, the expression $\mathbb{E}[R_t + d]$ is continuous in t and consequently the integral on the right-hand side may be interpreted as a Riemann integral. Therefore by the fundamental theorem of calculus and Assumption 3, we have for $t > t_0*$, where $t_0^* > t_0$ is such that $\frac{B}{t} \leq \frac{cE}{2 \ln t}$ for $t > t_0^*$, that

$$\frac{d}{dt} \mathbb{E}[R_t + d] \leq \int_s^t \left(\left(\frac{B}{u} - cT_u \right) \mathbb{E}[R_u + d] + B' T_u^{-1} \right) du \leq -\frac{cE}{2 \ln t} \mathbb{E}[R_t + d] + \frac{B'}{E} \ln t.$$

This yields for $t > t_0^*$,

$$\frac{d}{dt} \left(e^{\frac{cE}{2} \int_{t_0^*}^t (\ln s)^{-1} ds} \mathbb{E}[R_t + d] \right) \leq \frac{B'}{E} \ln t e^{\frac{cE}{2} \int_{t_0^*}^t (\ln s)^{-1} ds}$$

and

$$\begin{aligned} \mathbb{E}[R_t + d] &\leq \mathbb{E}[R_{t_0^*} + d] e^{-\frac{cE}{2} \int_{t_0^*}^t (\ln s)^{-1} ds} + \int_{t_0^*}^t \frac{B'}{E} \ln s e^{-\frac{cE}{2} \int_s^t (\ln u)^{-1} du} \\ &\leq \mathbb{E}[R_{t_0^*} + d] + \frac{B'}{E} \int_{t_0^*}^t \ln s e^{-\frac{cE}{2} \int_s^t (\ln u)^{-1} du} \\ &\leq \mathbb{E}[R_{t_0^*} + d] + \frac{B'}{E} \ln t \int_{t_0^*}^t e^{-\frac{cE}{2 \ln t}(t-s)} \\ &\leq \mathbb{E}[R_{t_0^*} + d] + B' \frac{2(\ln t)^2}{cE^2} \left(1 - e^{-\frac{cE}{2}(\ln t)^{-1}(t - t_0^*)} \right) \\ &\leq \mathbb{E}[R_{t_0^*} + d] + B' \frac{2(\ln t)^2}{cE^2}, \end{aligned}$$

where the first term on the right-hand side can be bounded via Proposition 2.4.1 and the inequalities in (2.4.23). $\square$

Corollary 2.4.4. *Under Assumption 1, 3, 4 and for $0 \leq \epsilon \leq \epsilon'$, the solution $(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon)$ to* (2.4.3) *is such that $\frac{\mathbb{E}[|X_t^\epsilon|^2+|Y_t^\epsilon|^2+|Z_t^\epsilon|^2]}{(\ln(e+t))^2}$ is bounded uniformly in time and in ϵ.*

Proof. By the lower bound on R in (2.4.23),

$$\mathbb{E}\left[|X_t^\epsilon|^2 + |Y_t^\epsilon|^2 + |Z_t^\epsilon|^2\right] \leq \mathbb{E}\left[\frac{R(X_t^\epsilon, Y_t^\epsilon, Z_t^\epsilon, T_t) + d}{a}\right],$$

which concludes by Lemma 2.4.3. $\square$

2.4.4 Form of Distorted Entropy

For $\epsilon \geq 0$, let $H^\epsilon(t)$ be the distorted entropy

$$\begin{aligned} H^\epsilon(t) \;=\; & \int \Bigg(\frac{\left|2\nabla_x h_t^\epsilon + 8S_0(\nabla_y h_t^\epsilon + \lambda^{-1}\nabla_z h_t^\epsilon)\right|^2}{h_t^\epsilon} \;+\; \frac{\left|\nabla_y h_t^\epsilon + S_1\lambda^{-1}\nabla_z h_t^\epsilon\right|^2}{h_t^\epsilon} \\ & +\; \beta(T_t^{-1}) h_t^\epsilon \ln(h_t^\epsilon)\Bigg) d\mu_{T_t}, \end{aligned} \tag{2.4.36}$$

where $S_0, S_1 > 0$ are the constants

$$S_0 := (1 + |D_x^2 U|_\infty^2)^{\frac{1}{2}}, \qquad S_1 := 2 + 28S_0^2 + 1024S_0^4 \tag{2.4.37}$$

and β is a second order polynomial (see (2.4.38) and the end of the proof for Proposition 2.4.6) to be determined by Proposition 2.4.6 and independent of ϵ.

Remark 2.4.2. This particular expression for H is not necessarily the best possible choice. However the above is a working expression and optimality is left as future work; see also [159].

The following auxiliary result can be found as Lemma 12 of [139]; its proof can also be found there.

Lemma 2.4.5. *For*

$$\Phi^*(h) = \frac{|M\nabla h|^2}{h},$$

where M is matrix-valued,

$$\Gamma_{L_t^{\epsilon*}, \Phi^*}(h) > \frac{(M\nabla h) \cdot [L_t^{\epsilon*}, M\nabla] h}{h}$$

holds for all $h \in \mathcal{C}_+^\infty$.

Notice the Φ^* appears in the first two terms of $H^\epsilon(t)$.
For the calculations below, recall the definitions (2.4.37) stated below for convenience:

$$\begin{aligned} S_0 &= (1+|D_x^2 U|_\infty^2)^{\frac{1}{2}}, \\ S_1 &= 2 + 28S_0^2 + 1024S_0^4. \end{aligned}$$

Using Lemma 2.4.5, the following proposition shows the distorted entropy (2.4.36) is a useful one.

Proposition 2.4.6. *There exist $\beta_0, \beta_1, \beta_2 > 0$ independent of ϵ such that for $\beta : \mathbb{R} \to \mathbb{R}$ given by*

$$\beta(x) := 1 + \beta_0 + \beta_1 x + \beta_2 x^2, \tag{2.4.38}$$

the operator Ψ_{T_t},

$$\Psi_{T_t}(h) \;:=\; \frac{\left|2\nabla_x h + 8S_0(\nabla_y h + \lambda^{-1}\nabla_z h)\right|^2}{h} + \frac{\left|\nabla_y h + S_1\lambda^{-1}\nabla_z h\right|^2}{h} + \beta(T_t^{-1})h\ln(h) \tag{2.4.39}$$

for $h \in \mathcal{C}_+^\infty$, satisfies

$$\Gamma_{L_t^{\epsilon *},\Psi_{T_t}}(h) \geq \frac{|\nabla h|^2}{h} \tag{2.4.40}$$

for all $0 \leq \epsilon \leq 1$.

Remark 2.4.3. β_0, β_1, β_2 depend on $\hat{\lambda}^2 := \max\left(|\lambda|^2, \left|\lambda^\top\right|^2, \left|\lambda^{-1}\right|^2, \left|\lambda^{-1}\right|\left|\lambda^\top\right|\right)$, $|D_x^2 U|_\infty$ and $|A|$. H satisfying property (2.4.40) is crucial for proving dissipation in Proposition 2.4.10.

Proof. Let Φ_1,Φ_2,Φ_3 be the terms in Ψ_{T_t},

$$\Phi_1(h) := \frac{\left|2\nabla_x h + 8S_0(\nabla_y h + \lambda^{-1}\nabla_z h)\right|^2}{h}, \tag{2.4.41a}$$

$$\Phi_2(h) := \frac{\left|\nabla_y h + S_1\lambda^{-1}\nabla_z h\right|^2}{h}, \tag{2.4.41b}$$

$$\Phi_3(h) := h\ln(h). \tag{2.4.41c}$$

Note that the Γ_Φ operator is linear in the Φ argument by linearity of $L_t^{\epsilon *}$, so that (2.4.40) can be written as $\Gamma_{L_t^{\epsilon *},\Phi_1}(h) + \Gamma_{L_t^{\epsilon *},\Phi_2}(h) + \beta(T_t^{-1})\Gamma_{L_t^{\epsilon *},\Phi_3}(h) \geq \frac{|\nabla h|^2}{h}$. Consider $\Gamma_{L_t^{\epsilon *},\Phi_3}$ first. Using the definition (2.4.6) of $\Gamma_{L_t^{\epsilon *},\Phi}$, the product and chain rule (2.4.8) and (2.4.7) for $L_t^{\epsilon *}$, and the coercivity property of A, we get

$$\begin{aligned}\Gamma_{L_t^{\epsilon *},\Phi_3}(h) &= \frac{1}{2}\bigg((\ln h+1)L_t^{\epsilon *}h + \frac{1}{h}\nabla h\cdot(A^\epsilon\nabla h) - (1+\ln h)L_t^{\epsilon *}h\bigg)\\ &= \frac{1}{2h}\nabla h\cdot(A^\epsilon\nabla h) \geq \frac{1}{2h}(\epsilon|\nabla_x h|^2 + \epsilon|\nabla_y h|^2 + A_c|\nabla_z h|^2). \end{aligned} \tag{2.4.42}$$

Since the goal is to show (2.4.40), the availability of (2.4.42) counteracts any negative contributions in the z-derivative term, and any order ϵ contributions in the x- and y-derivatives, from $\Gamma_{L_t^{\epsilon *},\Phi_1}$ and $\Gamma_{L_t^{\epsilon *},\Phi_2}$; this counterweight materialises as β.
For $\Gamma_{L_t^{\epsilon *},\Phi_1}$ and $\Gamma_{L_t^{\epsilon *},\Phi_2}$, $S_0 > 0$ and $S_1 > 0$ as in (2.4.37) are used. Beginning with Lemma 2.4.5, we have

$$\begin{aligned}&h\Gamma_{L_t^{\epsilon *},\Phi_2}(h)\\ &\quad > (\nabla_y + S_1\lambda^{-1}\nabla_z)h\cdot[L_t^{\epsilon *}, \nabla_y + S_1\lambda^{-1}\nabla_z]h\\ &\quad = (\nabla_y + S_1\lambda^{-1}\nabla_z)h\cdot(\nabla_x - \lambda^\top\nabla_z + \epsilon T_t^{-1}\nabla_y + S_1\nabla_y + S_1T_t^{-1}\lambda^{-1}A\nabla_z)h\\ &\quad = \nabla_x h\cdot\nabla_y h - \nabla_y h\cdot(\lambda^\top\nabla_z h) + \epsilon T_t^{-1}|\nabla_y h|^2 + S_1|\nabla_y h|^2 + S_1T_t^{-1}\nabla_y h\cdot(\lambda^{-1}A\nabla_z h)\\ &\qquad + S_1\nabla_x h\cdot(\lambda^{-1}\nabla_z h) - S_1(\lambda^{-1}\nabla_z h)\cdot(\lambda^\top\nabla_z h) + \epsilon T_t^{-1}S_1\nabla_y h\cdot(\lambda^{-1}\nabla_z h)\\ &\qquad + S_1^2\nabla_y h\cdot(\lambda^{-1}\nabla_z h) + S_1^2T_t^{-1}(\lambda^{-1}\nabla_z h)\cdot(\lambda^{-1}A\nabla_z h).\end{aligned}$$

In order to get a bound in terms of $(\partial_i h)^2$ terms rather than $\partial_i h\partial_j h$ terms, we bound the $\partial_i h\partial_j h$ terms in the following ways,

$$\begin{aligned}\nabla_x h\cdot\nabla_y h &\geq -\frac{1}{2}|\nabla_x h|^2 - \frac{1}{2}|\nabla_y h|^2,\\ -\nabla_y h\cdot(\lambda^\top\nabla_z h) &\geq -\frac{1}{6}|\nabla_y h|^2 - \frac{3}{2}\left|\lambda^\top\right|^2|\nabla_z h|^2,\\ S_1T_t^{-1}\nabla_y h\cdot(\lambda^{-1}A\nabla_z h) &\geq -\frac{1}{6}|\nabla_y h|^2 - \frac{3}{2}S_1^2T_t^{-2}\left|\lambda^{-1}\right|^2|A|^2|\nabla_z h|^2,\\ S_1\nabla_x h\cdot(\lambda^{-1}\nabla_z h) &\geq -\frac{1}{2}|\nabla_x h|^2 - \frac{1}{2}S_1^2\left|\lambda^{-1}\right|^2|\nabla_z h|^2,\\ \epsilon T_t^{-1}S_1\nabla_y h\cdot(\lambda^{-1}\nabla_z h) &\geq -\epsilon T_t^{-1}|\nabla_y h|^2 - \frac{\epsilon}{4}S_1^2T_t^{-1}\left|\lambda^{-1}\right|^2|\nabla_z h|^2,\\ S_1^2\nabla_y h\cdot(\lambda^{-1}\nabla_z h) &\geq -\frac{1}{6}|\nabla_y h|^2 - \frac{3}{2}S_1^4\left|\lambda^{-1}\right|^2|\nabla_z h|^2\end{aligned}$$

and using (2.4.3) gives

$$h\Gamma_{L_t^{\epsilon*},\Phi_2}(h) > -|\nabla_x h|^2 + (1 + 28S_0^2 + 1024S_0^4)|\nabla_y h|^2 \\ - \frac{1}{2}\hat{\lambda}^2\left(3 + 2S_1 + S_1^2 + 3S_1^4 + S_1^2 T_t^{-1}\left(|A| + \frac{\epsilon}{2}\right) + 3S_1^2 T_t^{-2}|A|^2\right)|\nabla_z h|^2. \tag{2.4.43}$$

The last term $\Gamma_{L_t^{\epsilon*},\Phi_1}$ compensates for the negative x-derivative. Again, beginning with Lemma 2.4.5, we have

$$\begin{aligned}
&h\Gamma_{L_t^{\epsilon*},\Phi_1}(h) \\
&\quad > (2\nabla_x + 8S_0(\nabla_y + \lambda^{-1}\nabla_z))h \cdot [L_t^{\epsilon*}, 2\nabla_x + 8S_0(\nabla_y + \lambda^{-1}\nabla_z)]h \\
&\quad = (2\nabla_x + 8S_0(\nabla_y + \lambda^{-1}\nabla_z))h \cdot (-2(D_x^2 U)(\nabla_y - \epsilon T_t^{-1}\nabla_x) \\
&\qquad + 8S_0(\nabla_x + \epsilon T_t^{-1}\nabla_y - \lambda^\top\nabla_z + \nabla_y + T_t^{-1}\lambda^{-1}A\nabla_z))h \\
&\quad = ((16S_0 I_n + 4\epsilon T_t^{-1}D_x^2 U)\nabla_x h) \cdot \nabla_x h + 2\nabla_x h \cdot ((-2D_x^2 U + 8S_0(1 + \epsilon T_t^{-1})I_n)\nabla_y h) \\
&\qquad + 2\nabla_x h \cdot (8S_0(-\lambda^\top + T_t^{-1}\lambda^{-1}A)\nabla_z h) + ((64S_0^2 I_n + 16S_0\epsilon T_t^{-1}D_x^2 U)\nabla_x h) \cdot \nabla_y h \\
&\qquad + 8S_0\nabla_y h \cdot ((-2D_x^2 U + 8S_0)\nabla_y h) + 8S_0\nabla_y h \cdot (8S_0(-\lambda^\top + T_t^{-1}\lambda^{-1}A)\nabla_z h) \\
&\qquad + ((64S_0^2 I_n + 16S_0\epsilon T_t^{-1}D_x^2 U)\nabla_x h) \cdot (\lambda^{-1}\nabla_z h) \\
&\qquad + ((-16S_0 D_x^2 U + 64S_0^2(1 + \epsilon T_t^{-1})I_n)\nabla_y h) \cdot (\lambda^{-1}\nabla_z h) \\
&\qquad + 64S_0^2(\lambda^{-1}\nabla_z h) \cdot ((-\lambda^\top + T_t^{-1}\lambda^{-1}A)\nabla_z h).
\end{aligned}$$

Bounding the $\partial_i h \partial_j h$ terms as for Φ_2, using (2.4.3) and (2.2.3) yields

$$
\begin{aligned}
&h\Gamma_{L_t^{\epsilon*},\Phi_1}(h) \\
&\quad > (16S_0 - 4\epsilon T_t^{-1}|D_x^2U|_\infty)|\nabla_x h|^2 \\
&\qquad - \Big(2|\nabla_x h|^2 + 2|D_x^2U|_\infty^2|\nabla_y h|^2 + 8(1+\epsilon T_t^{-1})|\nabla_x h|^2 + 8S_0^2(1+\epsilon T_t^{-1})|\nabla_y h|^2\Big) \\
&\qquad - \Big(2|\nabla_x h|^2 + 32S_0^2\hat{\lambda}^2\Big(1+T_t^{-2}|A|^2\Big)|\nabla_z h|^2\Big) \\
&\qquad - \Big((1+8S_0\epsilon T_t^{-1}|D_x^2U|_\infty^2)|\nabla_x h|^2 + (1024S_0^4 + 8S_0\epsilon T_t^{-1})|\nabla_y h|^2\Big) \\
&\qquad - \Big(16S_0|D_x^2U|_\infty|\nabla_y h|^2 - 64S_0^2|\nabla_y h|^2\Big) \\
&\qquad - \Big(32S_0^2|\nabla_y h|^2 + 32S_0^2\hat{\lambda}^2\Big(1+T_t^{-2}|A|^2\Big)|\nabla_z h|^2\Big) \\
&\qquad - \Big((1+8S_0\epsilon T_t^{-1}|D_x^2U|_\infty^2)|\nabla_x h|^2 + (1024S_0^4 + 8S_0\epsilon T_t^{-1})\hat{\lambda}^2|\nabla_z h|^2\Big) \\
&\qquad - \Big(\Big(2|D_x^2U|_\infty^2 + 32S_0^2(1+\epsilon T_t^{-1})\Big)|\nabla_y h|^2 + (32S_0^2 + 32\epsilon T_t^{-1}S_0^2)\hat{\lambda}^2|\nabla_z h|^2\Big) \\
&\qquad - 64S_0^2\hat{\lambda}^2\Big(1+T_t^{-2}|A|^2\Big)|\nabla_z h|^2 \\
&\quad \geq \Big(2 - 4(2+(1+4S_0^2)S_0)\epsilon T_t^{-1}\Big)|\nabla_x h|^2 \\
&\qquad + \Big(S_0^2(-28 - 1024S_0^2) - 8S_0(1+5S_0)\epsilon T_t^{-1}\Big)|\nabla_y h|^2 \\
&\qquad - \Big(S_0^2\hat{\lambda}^2(160 + 128T_t^{-2}|A|^2 + 1024S_0^2) + 8S_0\hat{\lambda}^2(1+4S_0)\epsilon T_t^{-1}\Big)|\nabla_z h|^2,
\end{aligned}
$$

so that

$$
\begin{aligned}
&h\Gamma_{L_t^{\epsilon*},\Phi_1}(h) \\
&\quad \geq \Big(2 - 4(2+(1+4S_0^2)S_0)\epsilon T_t^{-1}\Big)|\nabla_x h|^2 \\
&\qquad + \Big(S_0^2(-28 - 1024S_0^2) - 8S_0(1+5S_0)\epsilon T_t^{-1}\Big)|\nabla_y h|^2 \\
&\qquad - \Big(S_0^2\hat{\lambda}^2(160 + 128T_t^{-2}|A|^2 + 1024S_0^2) + 8S_0\hat{\lambda}^2(1+4S_0)\epsilon T_t^{-1}\Big)|\nabla_z h|^2.
\end{aligned}
$$

Matching powers in T_t^{-1} to take

$$
\begin{aligned}
\beta_0 &= \frac{1}{A_c}(S_0^2\hat{\lambda}^2(160 + 1024S_0^2) + \frac{1}{2}\hat{\lambda}^2(3 + 2S_1 + S_1^2 + 3S_1^4)) \\
\beta_1 &= \frac{1}{A_c}\left(4(2+(1+4S_0^2)S_0) + 8S_0(1+5S_0) + 8S_0\hat{\lambda}^2(1+4S_0) + \frac{1}{2}\hat{\lambda}^2\left(S_1^2\left(|A| + \frac{1}{2}\right)\right)\right) \\
\beta_2 &= \frac{1}{A_c}\left(128S_0^2\hat{\lambda}^2|A|^2 + \frac{3}{2}\hat{\lambda}^2S_1^2|A|^2\right),
\end{aligned}
$$

using $\epsilon \leq 1$ and putting together the bounds for $\Gamma_{L_t^{\epsilon*},\Phi_3}, \Gamma_{L_t^{\epsilon*},\Phi_2}, \Gamma_{L_t^{\epsilon*},\Phi_1}$ gives (2.4.40).

□

2.4.5 Log-Sobolev Inequality

Proposition 2.4.7. *Under Assumption 1, 2 and for $\epsilon \geq 0$, there exists constants $t_{ls}, A_* > 0$ and a finite order polynomial $r : (0, \infty) \to (0, \infty)$ with coefficients depending on U and λ but independent of ϵ such that the distorted entropy* (2.4.36) *satisfies*

$$H^\epsilon(t) \leq C_t \int \frac{|\nabla h_t^\epsilon|^2}{h_t^\epsilon} d\mu_{T_t}, \tag{2.4.44}$$

where for $t > t_{ls}$,

$$C_t = A_* + r\left(T_t^{-\frac{1}{2}}\right) e^{\hat{E} T_t^{-1}}. \tag{2.4.45}$$

Proof. Given Proposition 2.2.1, only the first two terms in the integrand of $H^\epsilon(t)$ are left, which lead directly to the inequality corresponding to A_*. □

2.4.6 Proof of Dissipation

Lemma 2.4.8 below constructs a sequence of compactly supported functions that are multiplied with the integrand in $H(t)$. It gives sufficient properties for retrieving a bound on $\partial_t H(t)$ after passing the derviative under the integral sign and passing the limit in the sequence of approximating initial densities. The key sufficient property turns out to be (2.4.46) below.
Let φ_k be given as in (2.4.1) and $\nu_k := \varphi_k * \mathbb{1}_{(-\infty, k^2]} \leq 1$ for $k > 0$.

Lemma 2.4.8. *For $k > 0$, define the smooth functions $\eta_k : \mathbb{R}^{2n+m+1} \to \mathbb{R}$*

$$\eta_k = \nu_k(-\ln(R + 2d)),$$

where $d > 0$ is the same as in (2.4.23)*. The following properties hold:*

1. *η_k is compactly supported;*
2. *η_k converges to 1 pointwise as $k \to \infty$;*
3. *for some constant $C > 0$ independent of k, t and $0 \leq \epsilon \leq \min(1, \epsilon')$*

$$L_t^\epsilon \eta_k \leq \frac{C T_t^{-1}}{k}. \tag{2.4.46}$$

Proof. By the quadratic assumption (2.2.7) on U and the bound (2.4.23) on R, R grows quadratically and in particular for an arbitrarily large constant $R_{(0)} > 0$, a compact set K can be chosen such that $R > R_{(0)}$ in $\mathbb{R}^{2n+m} \setminus K$; along with the support of ν_m being bounded below, the first statement is clear. The second statement is also trivial to check. The third statement is an application of (2.4.4), (2.4.5) and (2.4.24); by (2.4.4) and (2.4.5), we have

$$\begin{aligned}L_t^\epsilon \eta_m &= -\nu_m'(-\ln(R+2d))L_t^\epsilon \ln(R+2d)\\ &\quad + \nu_m''(-\ln(R+2d))(\nabla \ln(R+2d))^\top A^\epsilon \nabla \ln(R+2d).\end{aligned}$$

It can be seen that ν_m' and ν_m'' are estimated by terms at most of order m^{-1}; to see this, for all $x \in \mathbb{R}$,

$$\nu_m(x) = \int_{-\infty}^{m^2} \varphi_m(x-y)dy = \int_{x-m^2}^{\infty} \varphi_m(z)dz,$$

so that $0 \geq \nu_m'(x) = -\varphi_m(x-m^2) \geq -m^{-1}\max\varphi$ and $|\nu_m''(x)| = \left|\varphi_m'(x-m^2)\right| \leq m^{-2}\max\varphi'$. Therefore there exists a constant $\bar{C} > 0$ such that

$$\begin{aligned}L_t^\epsilon \eta_m &\leq -\nu_m'(-\ln(R+2d))\max(0, L_t^\epsilon \ln(R+2d))\\ &\quad + m^{-2}\max\varphi' \left|(\nabla \ln(R+2d))^\top A^\epsilon \nabla \ln(R+2d)\right|\\ &\leq \bar{C}\Big(m^{-1}\max(0, L_t^\epsilon \ln(R+2d)) + m^{-2}\left|(\nabla \ln(R+2d))^\top A^\epsilon \nabla \ln(R+2d)\right|\Big).\end{aligned}$$

A calculation using property (2.4.24) with (2.4.4) and (2.4.5) for L_t^ϵ reveals

$$\begin{aligned}L_t^\epsilon \ln(R+2d) &= \frac{L_t^\epsilon R}{R+2d} - \frac{(\nabla R)^\top A^\epsilon \nabla R}{(R+2d)^2}\\ &\leq \frac{-cT_tR + dT_t^{-1}}{R+2d} - \frac{\epsilon(|\nabla_x R|^2 + |\nabla_y R|^2) + A_c|\nabla_z R|^2}{(R+2d)^2}\\ &\leq \frac{-cT_t(R+d) + cT_td + dT_t^{-1}}{R+2d} - \frac{\epsilon(|\nabla_x R|^2 + |\nabla_y R|^2) + A_c|\nabla_z R|^2}{(R+2d)^2}\end{aligned}$$

and

$$(\nabla \ln(R+2d))^\top A^\epsilon \nabla \ln(R+2d) \leq (|A|+2)|\nabla \ln(R+2d)|^2 = (|A|+2)\left|\frac{\nabla R}{R+2d}\right|^2,$$

which are bounded above as claimed considering (2.4.23) and that ∇R grows linearly in space and is uniformly bounded in time.

□

Remark 2.4.4. Lemma 2.4.8 is different to Lemma 16 in [139]. We believe the first few equations in the proof of Lemma 16 in [139] contain a sign error; as a consequence the proofs in [139] beyond that point require significant modifications. Here we address this by modifying the truncation arguments we require, proving (2.4.46) instead of Lemma 17 of [139]. In addition, the finiteness of the distorted entropy is required, which is not a readily available result for our dynamics. This is the reason for using the perturbed dynamics in (2.4.3), so that the square integrability Theorem 7.4.1 in [20] can be used, which applies only to solutions of PDEs with uniformly elliptic operators. We prove the dissipation result for the original dynamics (2.1.3) by using a limiting argument.

For the convenience of the reader, we state here the corollary of Theorem 7.4.1 in [20] that will be used.

Theorem 2.4.9. *Under Assumptions 1, 3 and 4, for any $\epsilon, T > 0$, it holds that $m_t^\epsilon \in W^{1,1}(\mathbb{R}^n)$ for all $t > 0$ and*

$$\int_0^T \int \frac{|\nabla m_t^\epsilon(\zeta)|^2}{m_t^\epsilon(\zeta)} d\zeta dt < \infty.$$

The proof of Proposition 2.4.10 follows in the direction of Lemma 19 of [139].

Proposition 2.4.10. *Under Assumption 1, 2, 3 and 4 and for $0 < \epsilon \le \min(1, \epsilon')$, it holds that for any $0 < \alpha \le \frac{1}{2}(1 - \frac{\hat{E}}{E})$, there exists some constant $B > 0$ and some $t_H > 0$ both independent of ϵ, such that for all $t > t_H$,*

$$H^\epsilon(t) \le B\left(\frac{1}{t}\right)^{1-\frac{\hat{E}}{E}-2\alpha}. \tag{2.4.47}$$

Proof. Consider for $t \ge 0$ the auxiliary distorted entropies

$$\begin{aligned} H_k^\epsilon(t) &= \int \eta_k \Bigg(\frac{\left|2\nabla_x h_t^\epsilon + 8S_0(\nabla_y h_t^\epsilon + \lambda^{-1}\nabla_z h_t^\epsilon)\right|^2}{h_t^\epsilon} + \frac{\left|\nabla_y h_t^\epsilon + S_1\lambda^{-1}\nabla_z h_t^\epsilon\right|^2}{h_t^\epsilon} \\ &\qquad + \beta(T_t^{-1}) h_t^\epsilon \ln(h_t^\epsilon)\Bigg) d\mu_{T_t} \\ &= \int \eta_k(\Phi_1(h_t^\epsilon) + \Phi_2(h_t^\epsilon) + \beta(T_t^{-1})\Phi_3(h_t^\epsilon))d\mu_{T_t} = \int \eta_k \Psi_{T_t}(h_t^\epsilon) d\mu_{T_t}, \end{aligned} \tag{2.4.48}$$

where recall $h_t^\epsilon = m_t^\epsilon \mu_{T_t}^{-1}$, Φ_1, Φ_2, Φ_3 is as in (2.4.41) and η_k are as in Lemma 2.4.8. Due to the appearance of η_k, the function H_k^ϵ is differentiable and the order between the time derivative and the integral can be exchanged:

$$\frac{d}{dt} H_k^\epsilon(t) = \int \eta_k \partial_t(\Psi_{T_t}(h_t^\epsilon))d\mu_{T_t} + T_t' \int \eta_k \partial_{T_t}(\Psi_{T_t}(h_t^\epsilon)\mu_{T_t})dxdydz. \tag{2.4.49}$$

The terms will be considered separately. Since m_t^ϵ is the density of the law of (2.4.3) and $L_t^{\epsilon*}$ is the $L^2(\mu_{T_t})$ adjoint of L_t^ϵ, by Itô's rule for smooth compactly supported f

on $\mathbb{R}^{2n+m}$,

$$\int f\partial_t m_t^\epsilon = \partial_t \int f m_t^\epsilon = \int L_t^\epsilon f m_t^\epsilon = \int L_t^\epsilon f \frac{m_t^\epsilon}{\mu_{T_t}}\mu_{T_t} = \int f L_t^{\epsilon *}\left(\frac{m_t^\epsilon}{\mu_{T_t}}\right)\mu_{T_t}. \tag{2.4.50}$$

The first term in (2.4.49) is then bounded as follows.

$$\begin{aligned}\int \eta_k \partial_t(\Psi_{T_t}(h_t^\epsilon))d\mu_{T_t} &= \int \eta_k d\Psi_{T_t}(h_t^\epsilon).\partial_t h_t^\epsilon d\mu_{T_t} = \int \eta_k d\Psi_{T_t}(h_t^\epsilon).\frac{\partial_t m_t^\epsilon}{\mu_{T_t}}d\mu_{T_t}\\ &= \int \eta_k d\Psi_{T_t}(h_t^\epsilon).L_t^{\epsilon *}h_t^\epsilon d\mu_{T_t}\\ &= -\int 2\eta_k \Gamma_{L_t^{\epsilon *},\Psi_{T_t}}(h_t^\epsilon)d\mu_{T_t} + \int \eta_k L_t^{\epsilon *}(\Psi_{T_t}(h_t^\epsilon))d\mu_{T_t}\\ &= -\int 2\eta_k \Gamma_{L_t^{\epsilon *},\Psi_{T_t}}(h_t^\epsilon)d\mu_{T_t} + \int L_t^\epsilon \eta_k\Big(\Psi_{T_t}(h_t^\epsilon) + \beta(T_t^{-1})e^{-1}\Big)d\mu_{T_t}\\ &\le -2\int \eta_k \frac{|\nabla h_t^\epsilon|^2}{h_t^\epsilon}d\mu_{T_t} + \frac{CT_t^{-1}}{k}\int\Big(\Psi_{T_t}(h_t^\epsilon) + \beta(T_t^{-1})e^{-1}\Big)d\mu_{T_t},\end{aligned} \tag{2.4.51}$$

using Proposition 2.4.6 and Lemma 2.4.8, where $\beta(T_t^{-1})e^{-1}\int L_t^{\epsilon *}\eta_k d\mu_{T_t} = 0$ is added to force

$$\beta(T_t^{-1})(h_t^\epsilon \ln h_t^\epsilon + e^{-1}) \ge 0, \quad \text{so that} \quad \Psi_{T_t}(h_t^\epsilon) + \beta(T_t^{-1})e^{-1} \ge 0.$$

For the second term in (2.4.49), consider the Φ_1 and Φ_2 terms in the integrand

$$\eta_k \partial_{T_t}(\Psi_{T_t}\mu_{T_t}) = \eta_k \partial_{T_t}((\Phi_1 + \Phi_2 + \beta(T_t^{-1})\Phi_3)\mu_{T_t})$$

of $H_k(t)$ with the forms

$$\partial_{T_t}(\Phi_i(h_t^\epsilon)\mu_{T_t}) = \partial_{T_t}\left|M_i\nabla \ln\left(\frac{m_t^\epsilon}{\mu_{T_t}}\right)\right|^2 m_t^\epsilon, \qquad i = 1, 2$$

for the corresponding matrices M_1 and M_2 depending on S_0, S_1 and λ. Applying the partial derivative in T_t,

$$\partial_{T_t}(\Phi_i(h_t^\epsilon)\mu_{T_t}) = -2(M_i\nabla \ln h_t^\epsilon \cdot M_i\nabla\partial_{T_t}\ln \mu_{T_t})m_t^\epsilon, \tag{2.4.52}$$

and using definition (2.2.2) for μ_{T_t} and $Z_{T_t} = \int_{\mathbb{R}^{2n+m}} e^{-\frac{1}{T_t}\left(U(x)+\frac{|y|^2}{2}+\frac{|z|^2}{2}\right)} dxdydz$, gives

$$
\begin{aligned}
\partial_{T_t} \ln \mu_{T_t} &= \mu_{T_t}^{-1} \partial_{T_t} \left(Z_{T_t}^{-1} e^{-\frac{1}{T_t}\left(U(x)+\frac{|y|^2}{2}+\frac{|z|^2}{2}\right)} \right) \\
&= \mu_{T_t}^{-1} \left(- Z_{T_t}^{-2} \partial_{T_t} Z_{T_t} + \frac{Z_{T_t}^{-1}}{T_t^2} \left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right) \right) e^{-\frac{1}{T_t}\left(U(x)+\frac{|y|^2}{2}+\frac{|z|^2}{2}\right)} \\
&= \mu_{T_t}^{-1} \left(- \mu_{T_t} Z_{T_t}^{-1} \partial_{T_t} Z_{T_t} + \frac{\mu_{T_t}}{T_t^2} \left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right) \right) \\
&= - \int \frac{1}{T_t^2} \left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right) d\mu_{T_t} + \frac{1}{T_t^2} \left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right). \qquad (2.4.53)
\end{aligned}
$$

Note the exchange in differentiation and integration is justified by the bounds (2.2.7) on U. Integrating by parts in y and z (or simply using formulae for second moments) gives $\frac{n+m}{2T_t}$ for the $|y|^2$ and $|z|^2$ terms in the first integral. The integral over U can be dealt with using assumptions (2.2.4) and (2.2.5), to be specific:

$$
\begin{aligned}
\int U d\mu_{T_t} &\leq \int (a_M^2 |x|^2 + U_M) d\mu_{T_t} \leq \int \left(\frac{a_M^2}{r_1} (\nabla U \cdot x + U_g) + U_M \right) d\mu_{T_t} \\
&= \frac{a_M^2}{r_1} (nT_t + U_g) + U_M \\
\int U d\mu_{T_t} &\geq \int (a_m^2 |x|^2 + U_m) d\mu_{T_t} \geq \int \left(\frac{a_m^2}{r_2+1} (|\nabla U|^2 - U_g + |x|^2) + U_m \right) d\mu_{T_t} \\
&\geq \int \left(\frac{a_m^2}{r_2+1} (2\nabla U \cdot x - U_g) + U_m \right) d\mu_{T_t} = \frac{a_m^2}{r_2+1} (2nT_t - U_g) + U_m.
\end{aligned}
$$

Plugging into (2.4.53) gives

$$
p_1\left(T_t^{-1}\right) \leq \partial_{T_t} \ln \mu_{T_t} - \frac{1}{T_t^2} \left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} - \frac{n+m}{2} T_t \right) \leq p_2\left(T_t^{-1}\right). \qquad (2.4.54)
$$

where $p_1(x) = -\frac{a_M^2 n}{r_1} x - \left(\frac{a_M^2 U_g}{r_1} + U_M \right) x^2$ and $p_2(x) = -\frac{2a_m^2 n}{r_2+1} x + \left(\frac{a_m^2 U_g}{r_2+1} - U_m \right) x^2$. Substituting (2.4.53) back into (2.4.52),

$$
\begin{aligned}
\partial_{T_t} (\Phi_i(h_t^\epsilon) \mu_{T_t}) &\leq \left(|M_i \nabla \ln h_t^\epsilon|^2 + T_t^{-4} \left| M_i \nabla \left(U(x) + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right) \right|^2 \right) m_t^\epsilon \\
&\leq \Phi_i(h_t^\epsilon) \mu_{T_t} + \widetilde{C} T_t^{-4} \left(1 + |x|^2 + |y|^2 + |z|^2 \right) m_t^\epsilon \qquad (2.4.55)
\end{aligned}
$$

for a constant $\widetilde{C} \geq 0$ independent of k and ϵ by the quadratic assumption (2.2.5) on $|\nabla_x U|^2$ and $\eta_m \leq 1$.

For the last integrand in the last term of the right hand side of (2.4.49), namely the derivative over $\Phi_3(h_t^\epsilon)\mu_{T_t} = \frac{m_t^\epsilon}{\mu_{T_t}} \ln \frac{m_t^\epsilon}{\mu_{T_t}} \mu_{T_t}$, the left inequality of (2.4.54) gives

$$\begin{aligned}
&\partial_{T_t}(\beta(T_t^{-1})\Phi_3(h_t^\epsilon)\mu_{T_t}) \\
&= -T_t^{-2}\beta'(T_t^{-1})\Phi_3(h_t^\epsilon)\mu_{T_t} + \beta(T_t^{-1})\partial_{T_t} \ln \frac{m_t^\epsilon}{\mu_{T_t}} m_t^\epsilon \\
&= -T_t^{-2}\beta'(T_t^{-1})(\Phi_3(h_t^\epsilon) + e^{-1})\mu_{T_t} + T_t^{-2}\beta'(T_t^{-1})e^{-1}\mu_{T_t} - \beta(T_t^{-1})\partial_{T_t} \ln \mu_{T_t} m_t^\epsilon \\
&\leq T_t^{-2}\beta'(T_t^{-1})e^{-1}\mu_{T_t} + \beta(T_t^{-1}) \\
&\quad \cdot \left| p_1\left(T_t^{-1}\right) + \frac{1}{T_t^2}\left(-\frac{n+m}{2}T_t + U_M + a_M|x|^2 + \frac{|y|^2}{2} + \frac{|z|^2}{2} \right) \right| m_t^\epsilon, \qquad (2.4.56)
\end{aligned}$$

where in the last step $\Phi_3 + e^{-1} \geq 0$, $\beta_1, \beta_2 > 0$ and (2.2.7) have been used. Putting together the bounds (2.4.55) and (2.4.56) and applying Corollary 2.4.4 yields

$$\begin{aligned}
\int \eta_k \partial_{T_t}(\Psi_{T_t}(h_t^\epsilon)\mu_{T_t}) d\zeta &\leq q\left(T_t^{-1}\right)\left(H_k^\epsilon(t) + \mathbb{E}\left[1 + |X_t^\epsilon|^2 + |Y_t^\epsilon|^2 + |Z_t^\epsilon|^2\right]\right) \\
&\leq p\left(T_t^{-1}\right)\left(H_k^\epsilon(t) + \hat{C}\right), \qquad (2.4.57)
\end{aligned}$$

where p and q are some finite order polynomials with nonnegative coefficients, $\hat{C} > 0$, both independent of k and ϵ.

Returning to (2.4.49), collecting (2.4.51) and (2.4.57) then integrating from any $s \geq 0$ to $t > s$ gives

$$\begin{aligned}
H_k^\epsilon(t) - H_k^\epsilon(s) \leq 2\int_s^t \Bigg(&- \int \eta_k \frac{|\nabla h_u^\epsilon|^2}{h_u} d\mu_{T_u} + \frac{CT_u^{-1}}{k}(H^\epsilon(u) + \beta(T_u^{-1})e^{-1}) \\
&+ |T_u'| p\left(T_u^{-1}\right)\left(H_k^\epsilon(u) + \hat{C}\right)\Bigg) du. \qquad (2.4.58)
\end{aligned}$$

Fix an arbitrary $S > 0$. By Theorem 2.4.9, the log-Sobolev inequality (2.4.44), (2.2.5) and the finiteness of second moments (2.4.10), it holds that

$$\begin{aligned}
\int_0^S H^\epsilon(u) du &\leq \int_0^S C_u \int \frac{|\nabla h_u^\epsilon|^2}{h_u^\epsilon} d\mu_{T_u} du \\
&= \int_0^S C_u \int \frac{\left|\nabla m_u^\epsilon + T_u^{-1} m_u^\epsilon(\nabla_x U + y + z)\right|^2}{m_u^\epsilon} dx dy dz du < \infty. \qquad (2.4.59)
\end{aligned}$$

Then in (2.4.58) the $k \to \infty$ limit can be taken. Due to (2.4.59), the term denominated by k goes to zero. Applying Fatou's lemma (adding and subtracting $\beta(T_t^{-1})e^{-1} \int \eta_m d\mu_{T_t}$

wherever necessary for positivity) and using $\eta_m \le 1$, it holds that for $s < t$,

$$H^\epsilon(t) - H^\epsilon(s) \le -2\int_s^t \int \frac{|\nabla h_u^\epsilon|^2}{h_u^\epsilon} d\mu_{T_u} du + \int_s^t |T_u'| p\Big(T_u^{-1}\Big)\Big(H^\epsilon(u) + \hat{C}\Big) du \tag{2.4.60}$$

and for[6] $t_{ls} < s < t$,

$$H^\epsilon(t) - H^\epsilon(s) \le \int_s^t \bigg(\Big(|T_u'|p\Big(T_u^{-1}\Big) - 2C_u^{-1}\Big) H^\epsilon(u) + \hat{C}|T_u'|p\Big(T_u^{-1}\Big)\bigg) du. \tag{2.4.61}$$

Since $t^\alpha \gg (\ln t)^{\frac{\rho}{2}}$ for any $\rho, \alpha > 0$ and large enough $t > 0$, for any $\alpha > 0$, there exists $t_1 > \max(t_{ls}, t_0)$, where t_0 is as in Assumption 3, and $c_1, c_2 > 0$ independent of k, ϵ such that for all $t \ge t_1$,

$$|T_t'|p\Big(T_t^{-1}\Big) \le c_1 \left(\frac{1}{t}\right)^{1-\alpha}, \tag{2.4.62}$$

$$-2C_t^{-1} \le -c_2 \left(\frac{1}{t}\right)^{\frac{\hat{E}}{E}+\alpha}, \tag{2.4.63}$$

where the assumption $T_t \ge \frac{E}{\ln t}$ and (2.4.45) have been used. Using further that $E > \hat{E}$ by Assumption 3, then taking $\alpha < \frac{1}{2}(1 - \frac{\hat{E}}{E})$, there exists $t_2 \ge t_1$ independent of ϵ such that for $t \ge t_2$,

$$|T_t'|p\Big(T_t^{-1}\Big) - 2C_t^{-1} \le -c_3 \left(\frac{1}{t}\right)^{\frac{\hat{E}}{E}+\alpha} \tag{2.4.64}$$

and from (2.4.61), for $t_2 < s < t$,

$$H^\epsilon(t) - H^\epsilon(s) \le \int_s^t \left(-c_3 \left(\frac{1}{u}\right)^{\frac{\hat{E}}{E}+\alpha} H^\epsilon(u) + \hat{C} c_1 \left(\frac{1}{u}\right)^{1-\alpha}\right) du. \tag{2.4.65}$$

To obtain the corresponding differential inequality for all time, (2.4.65) can be divided

[6] t_{ls} from Proposition 2.4.7

by $t-s$, mollified with (2.4.1) for $0<k<1$ and the limit $s\to t$ can be taken:

$$\begin{aligned}
&\lim_{\hat\epsilon\to 0}\frac{1}{2\hat\epsilon}\int_{t-1}^{t+1}\varphi_k(t-u)(H^\epsilon(u+\hat\epsilon)-H^\epsilon(u-\hat\epsilon))du\\
&\quad\le \lim_{\hat\epsilon\to 0}\frac{1}{2\hat\epsilon}\int_{t-1}^{t+1}\varphi_k(t-u)\int_{u-\hat\epsilon}^{u+\hat\epsilon}\bigg(-c_3\Big(\frac{1}{u'}\Big)^{\frac{\hat E}{E}+\alpha}H^\epsilon(u')+\hat C c_1\Big(\frac{1}{u'}\Big)^{1-\alpha}\bigg)du'du\\
&\quad\le \int_{t-1}^{t+1}\varphi_k(t-u)\lim_{\hat\epsilon\to 0}\frac{1}{2\hat\epsilon}\int_{u-\hat\epsilon}^{u+\hat\epsilon}\bigg(-c_3\Big(\frac{1}{u'}\Big)^{\frac{\hat E}{E}+\alpha}H^\epsilon(u')+\hat C c_1\Big(\frac{1}{u'}\Big)^{1-\alpha}\bigg)du'du\\
&\quad= \int_{t-1}^{t+1}\varphi_k(t-u)\bigg(-c_3\Big(\frac{1}{u}\Big)^{\frac{\hat E}{E}+\alpha}H^\epsilon(u)+\hat C c_1\Big(\frac{1}{u}\Big)^{1-\alpha}\bigg)du
\end{aligned}$$

for $t\ge t_2+2$, where the second-to-last line follows from Fatou's lemma and dominated convergence (adding and subtracting $\beta(T_{u'}^{-1})e^{-1}$ to H^ϵ for Fatou); the last equality follows by the Lebesgue differentiation theorem. Therefore

$$\frac{d}{dt}(\varphi_k*H^\epsilon)(t)\le -c_3\Big(\frac{1}{t+1}\Big)^{\frac{\hat E}{E}+\alpha}(\varphi_k*H^\epsilon)(t)+\hat C'\Big(\frac{1}{t-1}\Big)^{1-\alpha}$$

for some constant $\hat C'>0$ independent of k,ϵ. Setting

$$\gamma_1(t):=c_3\Big(\frac{1}{t+1}\Big)^{\frac{\hat E}{E}+\alpha},\qquad \gamma_2(t):=\hat C'\Big(\frac{1}{t-1}\Big)^{1-\alpha}$$

and following the argument as per [139] from Lemma 6 in [138], there exists $t_3\ge t_2+2$, $c_4,c_5,c_6>0$ independent of k and ϵ such that for $t\ge t_3$,

$$\frac{d}{dt}\Big(\frac{\gamma_2}{\gamma_1}\Big)(t)=\frac{(t+1)^{\frac{\hat E}{E}+\alpha}}{(t-1)^{1-\alpha}}\Big(\frac{c_4}{t+1}-\frac{c_5}{t-1}\Big)\ge -c_6t^{-1},$$

so that there exists $t_4\ge t_3$ independent of k and ϵ such that for $t\ge t_4$,

$$\frac{d}{dt}\Big(\varphi_k*H^\epsilon-\frac{2\gamma_2}{\gamma_1}\Big)(t)\le -\gamma_1(t)\Big(\varphi_k*H^\epsilon(t)-\frac{2\gamma_2(t)}{\gamma_1(t)}\Big)$$

and consequently

$$\varphi_k*H^\epsilon(t)\le\frac{2\gamma_2(t)}{\gamma_1(t)}+\varphi_k*H^\epsilon(t_4)e^{-\int_{t_4}^t\gamma_1(u)du}. \tag{2.4.66}$$

Finally, from (2.4.65) (adding and subtracting $\beta(T_{u'}^{-1})e^{-1}$ to H^ϵ), it holds that for $t\ge$

$t_4 + 2$,

$$H^\epsilon(t) = \int_{t-2k}^{t} \varphi_k(t-k-s)ds H^\epsilon(t) \leq \int_{t-2k}^{t} \varphi_k(t-k-s)H^\epsilon(s)ds + \tilde{g}(2k) \tag{2.4.67}$$

for some $\tilde{g} : \mathbb{R} \to \mathbb{R}$ satisfying $\tilde{g}(k') \to 0$ as $k' \to 0$, so that (2.4.66) yields

$$H^\epsilon(t) \leq \frac{2\gamma_2(t-k)}{\gamma_1(t-k)} + \varphi_k * H^\epsilon(t_4)e^{-\int_{t_4}^{t-k}\gamma_1(u)du} + \tilde{g}(2k),$$

where $\varphi_k * H^\epsilon(t_4)$ can be bounded independently of k in a similar spirit to (2.4.67), and taking $k \to 0$ concludes the proof. □

Remark 2.4.5. The annealing schedule T_t is chosen to satisfy the relationship (2.4.64) between C_t^{-1} and $|T_t'|p\Big(T_t^{-1}\Big)$.

2.4.7 Degenerate noise limit

After taking advantage of the square integrability Theorem 7.4.1 in [20] for the case with a nondegenerate diffusion term in the proof of Proposition 2.4.10, the $\epsilon \to 0$ limit is taken to obtain the same dissipation inequality in this section.

Proof of Proposition 2.2.4. From (2.4.60), for any $0 \leq s < t$ and $0 < \epsilon \leq \min(1, \epsilon')$, it holds that

$$H^\epsilon(t) - H^\epsilon(s) \leq \int_s^t |T_u'|p\Big(T_u^{-1}\Big)\Big(H^\epsilon(u) + \hat{C}\Big)du,$$

where p is a finite order polynomial with nonnegative coefficients and $\hat{C} > 0$ is a constant both independent of ϵ. Therefore, mollifying in time and taking $s \to t$ as in the end of the proof by Proposition 2.4.10, it is straightforward that H^ϵ is uniformly bounded in[7] $0 \leq t \leq t_H$ and $0 < \epsilon \leq \min(1, \epsilon')$. Moreover by Proposition 2.4.10, the entropy $\int h_t^\epsilon \ln h_t^\epsilon d\mu_{T_t}$ is bounded uniformly in $t > t_H$ and $0 < \epsilon \leq \min(1, \epsilon')$. Therefore for any $t \geq 0$ by the de la Vallée-Poussin criterion (see for example [49]), the subset $\{h_t^\epsilon : 0 < \epsilon \leq \min(1, \epsilon')\} \subset L^1(\mu_{T_t})$ is uniformly integrable and consequently the Dunford-Pettis theorem implies the existence of a weak limit $g_t \in L^1(\mu_{T_t})$ for a (sub)sequence $(\epsilon_i)_{i\in\mathbb{N}}$ such that $\epsilon_i \to 0$,

$$h_t^{\epsilon_i} \rightharpoonup g_t, \quad \text{in } L^1(\mu_{T_t}) \quad \text{as } i \to \infty.$$

For any $S > 0$, any compactly supported smooth test function $\phi : [0, S) \times \mathbb{R}^{2n+m} \to \mathbb{R}$, omitting the dependence on the space variable $\zeta = (x, y, z)$ wherever convenient, denoting

[7] t_H from Proposition 2.4.10

$D_S := (0,S)\times\mathbb{R}^{2n+m}$ and using Itô's rule,

$$\begin{aligned}
0 &= \lim_{i\to\infty}\int_{D_S}(m_t^{\epsilon_i}-g_t\mu_{T_t})(-\partial_t-L_t)\phi dtd\zeta\\
&= \lim_{i\to\infty}\int_{D_S}\epsilon_i m_t^{\epsilon_i}(S_t^x+S_t^y)\phi dtd\zeta+\int_{D_S}g_t\mu_{T_t}(\partial_t+L_t)\phi dtd\zeta+\int_{\mathbb{R}^{2n+m}}m_0\phi(0,\zeta)dtd\zeta,\\
&= \int_{D_S}g_t\mu_{T_t}(\partial_t+L_t)\phi dtd\zeta+\int_{\mathbb{R}^{2n+m}}m_0\phi(0,\zeta)dtd\zeta, \qquad (2.4.68)
\end{aligned}$$

so that in the distributional sense of [20],

$$\begin{cases}\partial_t(g_t\mu_{T_t})=L_t^\top(g_t\mu_{T_t}) \quad \text{on } \mathbb{R}^{2n+m} \quad \forall t>0,\\ (g_0\mu_{T_0})=m_0.\end{cases} \qquad (2.4.69)$$

By Proposition 2.4.1, the solution to (2.4.69) is unique in the class of integrable solutions and since m_t belongs in this same class, it holds that

$$g_t\mu_{T_t}=m_t$$

for all $t\in[0,S]$, which is that

$$m_t^{\epsilon_i}\rightharpoonup m_t, \quad \text{in } L^1(\mu_{T_t}) \quad \text{as } i\to\infty.$$

for all $0\leq t<S$. By Corollary 3.8 in [25], there exists a sequence $(\hat{m}_t^i)_{i\in\mathbb{N}}$ made up of convex combinations of $m_t^{\epsilon_i}$ that converge strongly to m_t in L^1, hence a subsequence $(\hat{m}_t^{i_j})_{j\in\mathbb{N}}$ that convergences pointwise almost everywhere. By Fatou's lemma, convexity of $f(x)=x\ln x\geq e^{-1}$ for $x>0$ and Proposition 2.4.10, for $t>t_H$, we get

$$\int h_t\ln h_t d\mu_{T_t}=\int m_t\ln\left(\frac{m_t}{\mu_{T_t}}\right)\leq\liminf_{j\to\infty}\int\hat{m}_t^{i_j}\ln\left(\frac{\hat{m}_t^{i_j}}{\mu_{T_t}}\right)\leq B\left(\frac{1}{t}\right)^{1-\frac{\hat{E}}{E}-2\alpha}.$$

□

2.5 Some additional results

Before concluding, we complement our results about convergence to the global minimum with the analog of Proposition 2.4.10 for the $T_t=T>0$ sampling case and a result about the choice of the annealing schedule.
For completeness, we also show exponential convergence to equilibrium for the generalised Langevin equation (2.1.3) with constant temperature. Part of the analysis used in the proof of Proposition 2.4.10 can be used for the sampling case and $T_t=T$, i.e. working

only with the partial time derivatives mentioned above for the invariant distribution.

Proposition 2.5.1. *Let Assumption 1 and 4 hold and let $T_t = T$ for all t for some constant $T > 0$. There exist constants $C^c, C_* > 0$ such that*

$$\int |h_t - 1| d\mu_T \leq C^c e^{-\frac{C_*^{-1}}{2}t},$$

for all $t > 0$.

Proof. After Pinsker's inequality (2.2.12) and consideration of the definition (2.4.36) of H, what remains is the partial time derivative part of the proof of Proposition 2.4.10. The proof concludes by the same calculations as in Proposition 2.4.10, keeping in mind $T_t' = 0$, until (2.4.61) followed by the Grönwall argument. Note that (2.4.45) and (2.4.63) are not required and a log-Sobolev constant (in t also) works, in which case (2.4.44) and hence the current argument follow without requiring Assumption 2. The limiting ϵ argument as in Proposition 2.2.4 is the same. $\square$

Proposition 2.5.2. *Under Assumption 1, 3 and 4, the schedule $T_t = \frac{E}{\ln(e+t)}$, $E > \hat{E}$ is optimal in the sense that for any differentiable $f : \mathbb{R}_+ \to \mathbb{R}_+$, if*

$$T_t = \frac{1}{f(t)}\left(\frac{\hat{E}}{\ln(e+t)}\right), \tag{2.5.1}$$

C_t is the log-Sobolev factor (2.4.45) *and p is the finite order polynomial with nonnegative coefficients from the proof of Proposition 2.4.10, then the relation*

$$2C_t^{-1} \gg |T_t'| p\left(T_t^{-1}\right) \tag{2.5.2}$$

holds for large times only if $\limsup_{t\to\infty} f(t) \leq 1$.

Proof. Suppose there exists a constant $\delta > 0$ and times $(t_i)_{i\in\mathbb{N}}$ such that $0 < t_i \to \infty$ and

$$f(t_i) \geq 1 + \delta \quad \forall i.$$

From (2.4.45),

$$C_t^{-1} \sim \mathcal{O}(e^{-\hat{E}T_t^{-1}} T_t^{-1}),$$

which after substituting in (2.5.1) gives

$$e^{-\hat{E}T_t^{-1}} T_t^{-1} = (e+t)^{-f(t)} \frac{f(t)\ln(e+t)}{\hat{E}} \sim \mathcal{O}(t^{-f(t)} f(t) \ln t). \tag{2.5.3}$$

Compare this to

$$|T_t'|p\Big(T_t^{-1}\Big) \propto \frac{p(f(t)\ln(e+t))}{(f(t)\ln(e+t))^2}\bigg(\frac{f(t)}{e+t} + |f'(t)|\ln(e+t)\bigg), \tag{2.5.4}$$

which has order at least $(tf(t))^{-1}(\ln t)^{-2}$. For $t = t_i$, i large enough, $f(t) \geq 1+\delta$ and so

$$t^{-f(t)}f(t)\ln t \ll (tf(t))^{-1}(\ln t)^{-2}, \tag{2.5.5}$$

which violates (2.5.2). □

Remark 2.5.1. One can strengthen the proposition by making precise the form of p from Proposition 2.4.10, which will determine how slowly $f(t)$ is allowed to converge to 1; in fact p should be at least sixth order. This seems inconsequential with respect to optimality and so is omitted.

2.6 Conclusions

We explored the possibility of using the generalised Langevin equations in the context of simulated annealing. Our main purpose was to establish convergence as for the underdamped Langevin equation and provide a proof of concept in terms of performance improvement. Although the theoretical results hold for any scaling matrix A given the stated restrictions, we saw in our numerical results that its choice has great impact on the performance. In Section 2.3, A_2, A_3 or A_4 seemed to improve the exploration on the state space and/or the success proportion of the algorithm. There is plenty of work still required in terms of providing a more complete methodology for choosing A. This is left as future work and is also closely linked with time discretisation issues as a poor choice for A could lead to numerical integration stiffness. This motivates the development and study of improved numerical integration schemes, in particular, the extension of the conception and analysis on numerical schemes such as BAOAB [119] for the Langevin equation for (2.1.3) and the extension of the work in [143] for non-identity matrices λ and A. See [121] for work in this direction.

In addition, the system in (2.1.3) is not the only way to add an auxiliary variable to the underdamped Langevin equations in (2.1.2) whilst retaining the appropriate equilibrium distribution. Our choice was motivated by a clear connection to the generalised Langevin equation (2.1.4) and link with accelerated gradient descent, but it could be the case that a different third or higher order equations could be used with possibly improved performance. Along these lines, one could consider adding skew-symmetric terms as in [56]. As regards to theory, an interesting extension could involve establishing how the results here can be extended to establish a comparison of optimisation and sampling in a nonconvex setting for an arbitrary number of dimensions similar to [130]. We leave for future work finding optimal constants in the convergence results, investigating dependence on parameters and how the limits of these parameters and constants relate to existing results for the Langevin equation in (2.1.2) in [139, 159]. Finally, one could also aim to extend large deviation results in [107, 131, 169] for the overdamped Langevin dynamics to the underdamped and generalised case.

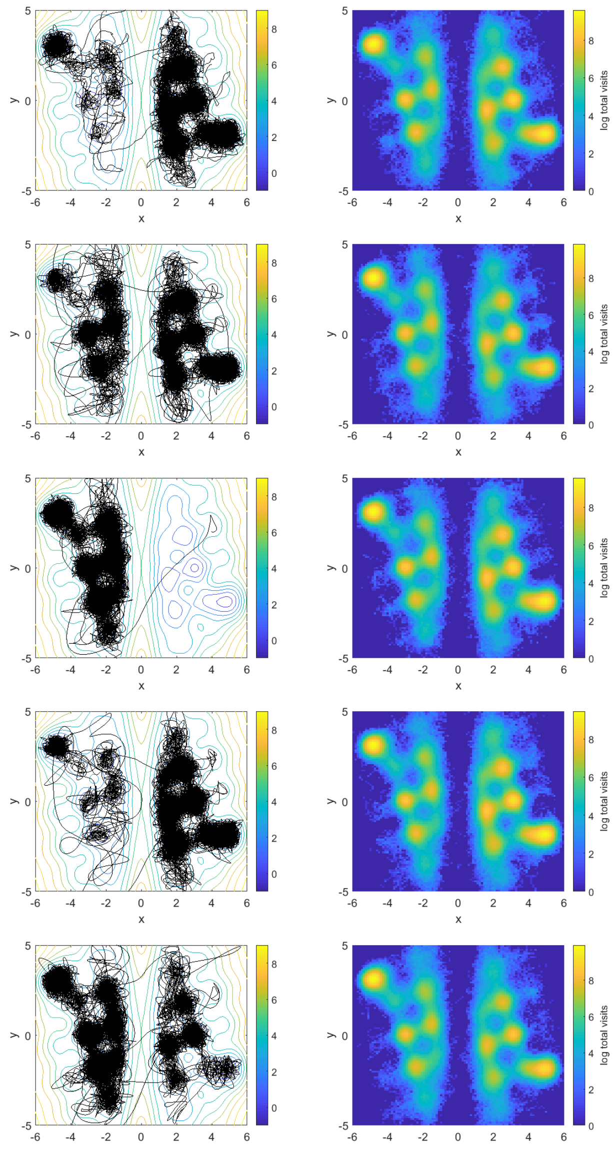

Fig. 2.3.1: Dynamics in order from top: (2.3.2), (2.3.1) with $A = A_1, \ldots, A_4$. Left: One instance of noise realisation. Right: Log histogram of 20 independent runs. See Section 2.3.2 for comments.

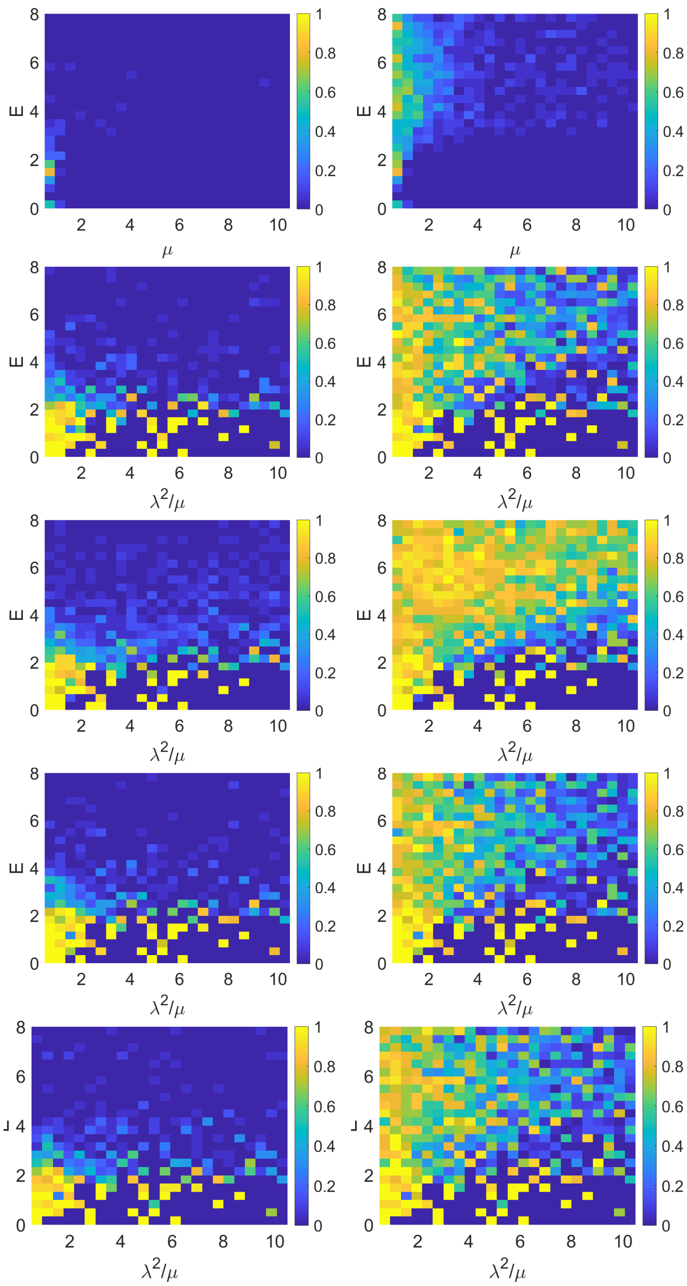

Fig. 2.3.2: Proportion of simulations close to the global minimum for the Alpine function as U. Panels from top to bottom: Langevin (2.3.2), generalised Langevin (2.3.1) with $A = A_1, A_2, A_3, A_4$. Left: Final position. Right: time-average of last 5000 iterations. We use $\gamma = 3$ for improving colour contrast, plots are visually similar for $\gamma = 1$.

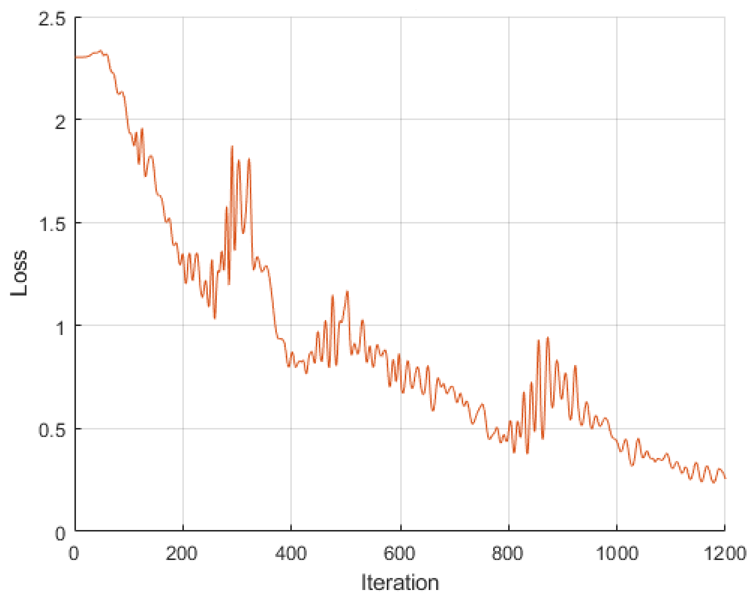

Fig. 2.3.3: Cross entropy between prediction and target over iterations of (2.3.1).

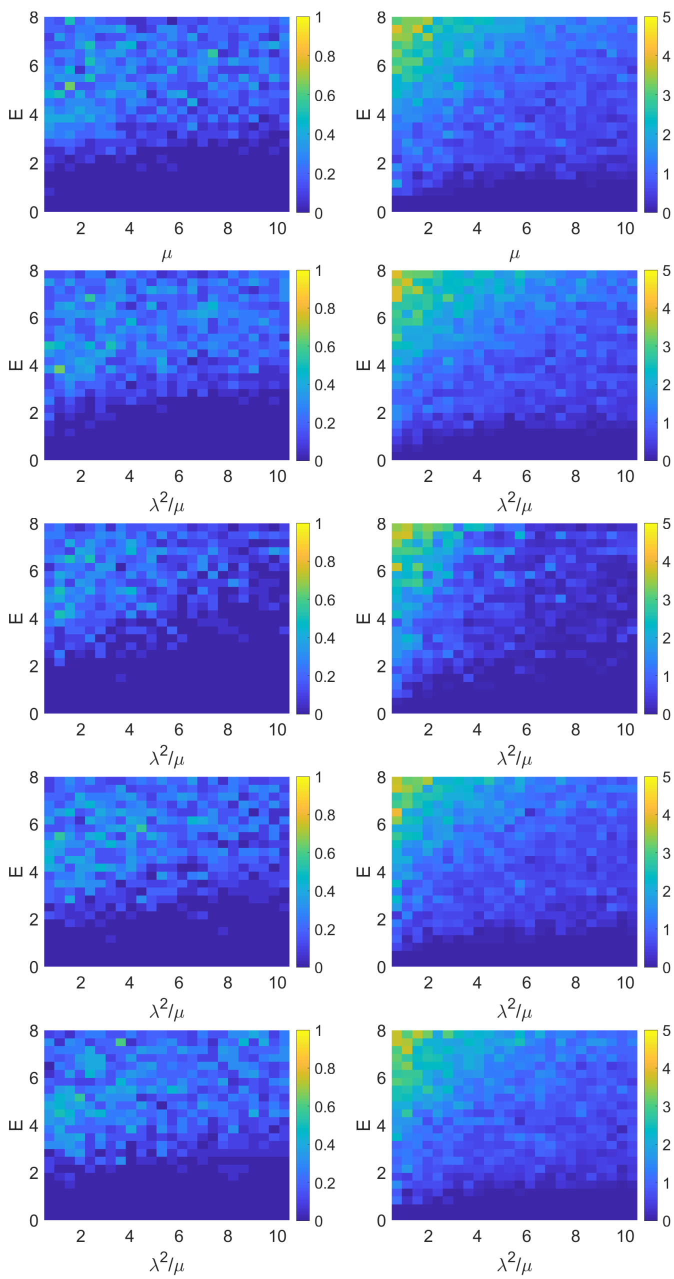

Fig. 2.3.4: Both proportion of success and numerical transition rates for $U = U_2$. Panels from top to bottom: (2.3.2), (2.3.1) with $A = A_1, A_2, A_3, A_4$. Left: Proportion satisfying the optimality tolerance for time-average of last 5000 iterations. Right: Average number of crossings of position averages over 5000 iterations across $\{x_1 = 0\}$ for each independent run. The remaining details are as in caption of Figure 2.3.2.

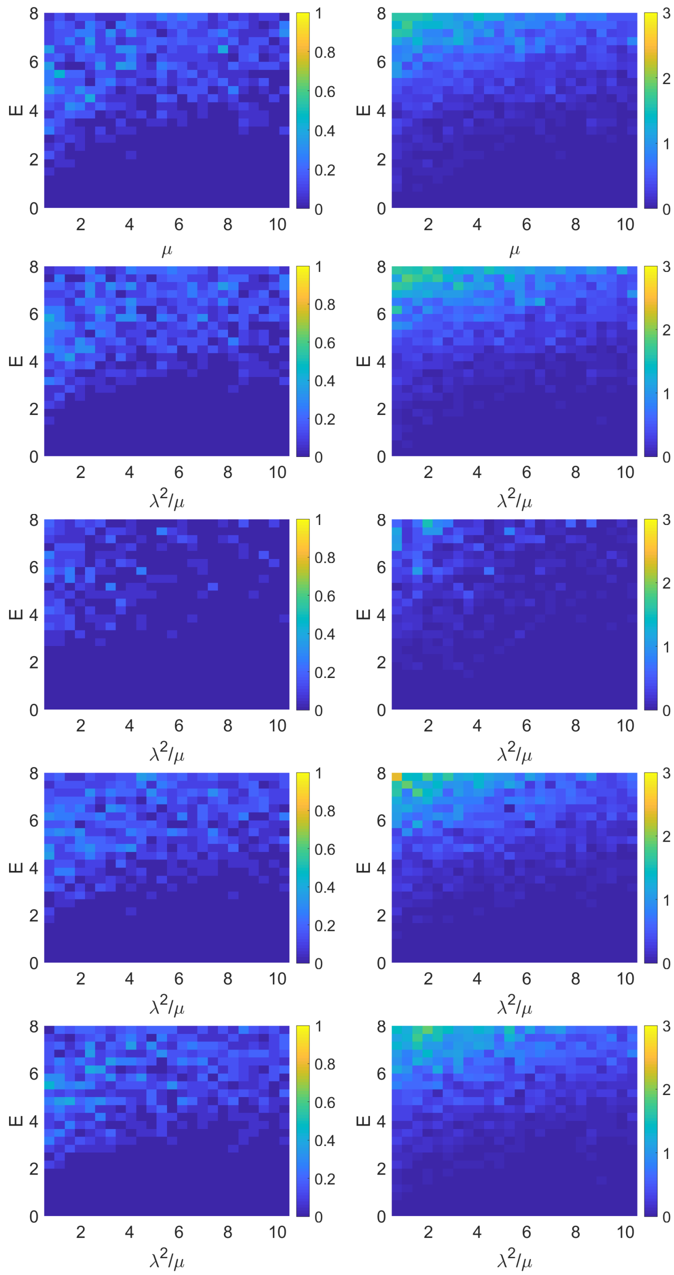

Fig. 2.3.5: Results for $U = U_3$. Details are as in caption of Figure 2.3.4.

3

Optimal friction for underdamped Langevin sampling

The contents of this chapter are from the paper [36] written in collaboration with N. Kantas, T. Lelièvre and G. Pavliotis.

3.1 Introduction

Let π be a probability measure on $\mathbb{R}^n$ with smooth positive bounded density, also denoted π, with respect to the Lebesgue measure on $\mathbb{R}^n$ and let $f \in L^2(\pi)$ be an observable or test function. In a range of applications including molecular dynamics [23, 120, 124] and machine learning [146, 191, 193], a quantity of interest is the expectation of f with respect to π,

$$\pi(f) := \int f d\pi,$$

which is analytically intractable and is numerically approximated most commonly by Markov Chain Monte Carlo (MCMC) methods, whereby π is sampled by simulating an ergodic Markov chain $(X_k)_{1\leq k\leq N}$ with π as its unique invariant measure and $\pi(f)$ is approximated by the empirical average $\frac{1}{N}\sum_{k=1}^{N} f(X_k)$. MCMC methods enjoy central limit theorems for many Markov chains employed, the most well-known (class) of such methods being the Metropolis-Hastings algorithm [89, 136]. Recent efforts have been to develop MCMC methods suited to settings where $n \gg 1$ and where point evaluations of π or its gradients are computationally expensive; these methods include slice sampling [54, 147], Hamiltonian Monte Carlo [14, 53, 148], piecewise-deterministic Markov processes [16, 24, 189] and those based on discretisations of continuous-time stochastic dynamics [60, 124, 129] together with divide-and-conquer and subsampling approaches [8].

In this chapter, the underdamped Langevin dynamics (1.0.1) are considered as noted in the introduction of the thesis, with the generality of a friction matrix. Denoting $\mathbb{S}^n_{++}$ as the set of real symmetric $n \times n$ positive definite matrices, the underdamped Langevin dynamics with mass $M \in \mathbb{S}^n_{++}$ and friction matrix $\Gamma \in \mathbb{S}^n_{++}$ is given by the $\mathbb{R}^{2n}$-valued solution (q_t, p_t) to

$$dq_t = M^{-1} p_t dt \tag{3.1.1a}$$

$$dp_t = -\nabla U(q_t) - \Gamma M^{-1} p_t dt + \sqrt{2\Gamma} dW_t, \tag{3.1.1b}$$

where $\sqrt{\Gamma} \in \mathbb{R}^{n\times n}$ is any matrix satisfying

$$\sqrt{\Gamma}\sqrt{\Gamma}^\top = \Gamma,$$

the function $U : \mathbb{R}^n \to \mathbb{R}$ is the associated smooth potential or negative log density such that $\pi \propto e^{-U}$ and W_t denotes a standard Wiener process on $\mathbb{R}^n$. The probability distribution from underdamped Langevin dynamics converges under general assumptions to the invariant probability measure given by

$$\tilde{\pi}(dq, dp) = Z^{-1} e^{-U(q) - \frac{p^\top M^{-1} p}{2}} dq dp \tag{3.1.2}$$

for a normalising constant Z and there have been numerous recent works [41, 47, 57, 65, 90, 117, 141, 173] on its discretisations in terms of the quality of convergence to $\tilde{\pi}$ over time; in this chapter, the goal is to optimise $\Gamma \in \mathbb{S}^n_{++}$ directly with respect to the asymptotic variance in the convergence of

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

to $\pi(f)$ for any particular f (or a finite set of observables) depending on q as $T \to \infty$. We mention that parameter tuning in MCMC methods is a widely considered topic [4, 196] (and references within). Specifically for underdamped Langevin dynamics, tuning the momentum part of $\tilde{\pi}$ with respect to reducing metastability or computational effort was considered in [166, 181, 188]. The choice of friction (as a scalar) has been a subject of consideration as early as in [96], then in [3, 26, 110, 178] within the context of molecular dynamics and also in [47, 56]. Most of these works make use of different measures for efficiency. The present work constitutes the first systematic gradient procedure for choosing the friction matrix in an optimal manner, with respect to an appropriate cost criterion.

3.1.1 Outline of approach

Complementary to the introduction of the thesis, we give here a concise description of our approach, precise statements can be found in the main Theorems 3.3.1 and 3.3.3. It is known using results from [172] and [15] that, under suitable assumptions on U and f, a central limit theorem

$$\frac{1}{\sqrt{T}}\int_0^T (f(q_t)-\pi(f))dt \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2) \qquad \text{as } T\to\infty \tag{3.1.3}$$

holds and that σ^2, the asymptotic variance, has the form

$$\sigma^2 = 2\int \phi(f-\pi(f))d\tilde{\pi} \tag{3.1.4}$$

where ϕ is a solution to the Poisson equation

$$-L\phi = f-\pi(f) \tag{3.1.5}$$

and L denotes the infinitesimal generator associated to (3.1.1). Two key observations are then made. Firstly, at any $\Gamma\in\mathbb{S}^n_{++}$ and for any direction $\delta\Gamma\in\mathbb{R}^{n\times n}$ in the friction matrix, the directional derivative of σ^2 evaluated at Γ in the direction $\delta\Gamma$, denoted $d\sigma^2.\delta\Gamma$, is given by the formula

$$d\sigma^2.\delta\Gamma = -2\int(\nabla_p\phi)^\top\delta\Gamma\nabla_p\tilde{\phi}d\tilde{\pi}, \tag{3.1.6}$$

where ϕ is the solution to (3.1.5) at Γ and $\tilde{\phi}$ is given by

$$\tilde{\phi}(q,p)=\phi(q,-p). \tag{3.1.7}$$

A direction $\delta\Gamma$ that guarantees a decrease in σ^2 is then

$$\Delta\Gamma := \int\nabla_p\phi\otimes\nabla_p\tilde{\phi}d\tilde{\pi} \tag{3.1.8}$$

where $\otimes$ denotes the outer product. Similarly, taking $\delta\Gamma$ to be the diagonal elements of (3.1.8) or $\delta\Gamma = I_n\int\nabla_p\phi\cdot\nabla_p\tilde{\phi}d\tilde{\pi}$ give in both cases a negative change in asymptotic variance respectively for diagonal Γ and Γ of the form cI_n. The second observation is that since the solution ϕ to the Poisson equation (3.1.5) is known to be given by

$$\phi(q,p)=\int_0^\infty\mathbb{E}[f(q_t)]dt, \tag{3.1.9}$$

where (q_t,p_t) solves (3.1.1) with initial condition $(q_0,p_0)=(q,p)$, given convexity conditions on the potential U and under suitable assumptions, we have

$$\nabla_p\phi=\int_0^\infty\mathbb{E}[\nabla f(q_t)^\top D_p q_t]dt, \tag{3.1.10}$$

where $D_p q_t$ denotes the $\mathbb{R}^{n\times n}$-matrix made of partial derivatives of q_t with respect to the initial condition p in momentum. In this case, $D_p q_t$ satisfy the dynamics that result from taking partial derivatives in (3.1.1), which are susceptible to algorithmic simulation. Moreover, under strong enough convexity assumptions on U, the process also decays to zero exponentially quickly, so that the infinite time integral (3.1.10) can be accurately approximated with a truncation using short simulations of $D_p q_t$ for adaptive estimations of the direction (3.1.8) in Γ. This leads to an adaptive algorithm involving the selection of Γ in an appropriate constrained set, of which we illustrate the performance with numerical examples.

Examples where improved Γ can be found analytically are presented in Section 3.4. Numerical illustrations making use of (3.1.6) and (3.1.10) are presented in Sections 3.5. In particular, the algorithm is applied on the problem of finding the posterior mean in a Bayesian logistic regression inference problem for two datasets with hundreds of dimensions, where the best friction matrices found in both cases are close to zero (for example $\Gamma = 0.1 I_n$ performs well compared to $\Gamma = I_n$, demonstrating reduced variance of almost an order of magnitude in Tables 3.5.2 and 3.5.3).
To use the asymptotic variance for a particular observable (or a set of them) and to use measures for the quality of convergence to $\tilde{\pi}$ or to minimise an autocorrelation time as considered in [3, 26, 47, 96, 110, 178] can be conflicting goals. To elaborate, in [96], the autocorrelation time was used as the point of comparison in the Gaussian target measure case for the optimal friction. For $n = 1$, $\omega, \gamma > 0$, $U(q) = \frac{1}{2}\omega^2 q^2$, $M = 1$, $\Gamma = \gamma$, the autocorrelation functions for (3.1.1) satisfies

$$\frac{d}{dt}\begin{pmatrix}\mathbb{E}(q_t q_0)\\ \mathbb{E}(p_t q_0)\end{pmatrix} = \begin{pmatrix}0 & 1\\ -\omega^2 & -\gamma\end{pmatrix}\begin{pmatrix}\mathbb{E}(q_t q_0)\\ \mathbb{E}(p_t q_0)\end{pmatrix}. \tag{3.1.11}$$

By considering the eigenvalues, the conclusion in [97] is that the optimal γ for minimising the magnitude of $\mathbb{E}(q_t q_0)$ is given by the critical damping $\gamma = 2\omega$, see Figure 3.1.1. A similar conclusion can be made when considering the spectral gap[158]. On the other hand, if $f(q) = q$ in our setting, formally, the quantity $\sigma^2 = 2\int\int_0^\infty \mathbb{E}(q_t q_0) dt d\tilde{\pi}(q_0, p_0)$ is the asymptotic variance due to (3.1.4) and (3.1.9). Despite the appearance of $\mathbb{E}(q_t q_0)$ as before, Corollary 3.4.8 asserts that $\gamma = 0$ is optimal for the asymptotic variance. A more detailed discussion about Corollary 3.4.8 is given in Section 3.4.2. This difference emphasizes that, at the cost of generic convergence to $\tilde{\pi}$, the tuning of Γ here is directed at variance reduction for a particular observable, in this case $f(q) = q$. However, multiple asymptotic variances can be used for the objective function to minimise, so that Γ can be optimised with respect to several observables of interest simultaneously. Remark 3.5.1

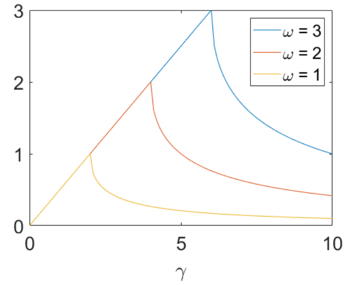

Fig. 3.1.1: The values $\min_i(|\mathrm{Re}(\lambda_i)|)$, where λ_i are the eigenvalues of the matrix appearing in the square matrix on the right-hand side of (3.1.11), also the spectral gap for the generator of (3.1.1) with $n = 1$, $\omega > 0$, harmonic potential $U(q) = \frac{1}{2}\omega^2 q^2$, $M = 1$ and $\Gamma = \gamma$. Critical values of γ are given by 2ω.

describes the implementation for a linear combination of asymptotic variances at no extra cost in terms of evaluations of π or its gradients.

The rest of the chapter is organised as follows. In Section 3.2, we provide a mathematical setting in which the underdamped Langevin dynamics with a friction matrix and in particular (3.1.1) has a well-defined solution and satisfies the central limit theorem for suitable observables, together with notations used throughout the chapter. In Section 3.3, prerequisite results and the main formulae (3.1.6) and (3.1.10) are precisely stated. Exact results concerning improvements in Γ including the quadratic U, quadratic f and linear f cases are given in Section 3.4. Numerical methods in approximating (3.1.8) together with an algorithm resulting from (3.1.6) and (3.1.10) is outlined and detailed in Algorithm 1 and 2 respectively in Section 3.5, alongside examples of U and f where improvements in variance are observed. In Section 3.6, deferred proofs are given. In Section 3.7, we conclude and discuss future work.

3.2 Setting

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, $(\mathcal{F}_t)_{t\geq 0}$ be a normal (satisfying the usual conditions) filtration with $(W_t)_{t\geq 0}$ a standard Wiener process on $\mathbb{R}^n$ with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ and let $\tilde{\pi}$ be a probability measure given by (3.1.2) for some potential function $U : \mathbb{R}^n \to \mathbb{R}$ and mass matrix $M \in \mathbb{S}^n_{++}$.

The set of smooth compactly supported functions is denoted C_c^∞. Following the notation of [62], we denote the infinitesimal generator (see (3.6.5) for a definition) associated to (3.1.1) as L, which is given formally by its differential operator form, denoted $\mathcal{L}$, when acting on the subset $C_c^\infty(\mathbb{R}^{2n})$,

$$\mathcal{L} = p^\top M^{-1}\nabla_q - \nabla U(q)^\top \nabla_p - p^\top M^{-1}\Gamma\nabla_p + \nabla_p^\top \Gamma \nabla_p. \tag{3.2.1}$$

Its formal $L^2(\mathbb{R}^n)$-adjoint $\mathcal{L}^\top$ satisfies

$$\mathcal{L}^\top \tilde{\pi} = 0, \tag{3.2.2}$$

so that $\tilde{\pi}$ (see (3.1.2)) is an invariant probability measure for (3.1.1) for a normalisation constant Z. Let

$$L_0^2(\pi) := \{g \in L^2(\pi) : \int g d\pi = 0\}$$

and similar for $L_0^2(\tilde{\pi})$. The notation D^2U will be used for the Hessian matrix of U. As in the introduction, $I_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix. For a matrix A, $|A|$ denotes the operator norm associated with the Euclidean norm. e_i is used to denote the i^{th} Euclidean basis vector. For $A, B \in \mathbb{R}^{n\times n}$, $A : B := \sum_{i,j} A_{ij}B_{ij}$ and $A_S = \frac{1}{2}(A + A^\top)$. $\langle \cdot, \cdot \rangle_{\tilde{\pi}}$ denotes the inner product in $L^2(\tilde{\pi})$ and similar for π.

3.2.1 Semigroup bound, Poisson equation and central limit theorem

In this section, a central limit theorem for the solution to (3.1.1) is established, where the resulting asymptotic variance will be used as a cost function to optimise Γ with respect to. Specifically, it will be shown that under some weighted L^∞ bound on the observable $f \in L^2(\pi)$, the estimator π_T for the unique solution (q_t, p_t) to (3.1.1) converges to $\pi(f)$ as $T \to \infty$ such that (3.1.3) holds with (3.1.4).

It is well known that the asymptotic variance can be expressed in terms of the solution to the Poisson equation (3.1.5) using the Kipnis-Varadhan framework, see for example Chapter 2 in [113], Section 3.1.3 in [124], [29] and references therein. In order to show that the expression (3.1.9) is indeed a solution to the Poisson equation (3.1.5), the exponential decay of the semigroup (see (3.6.4)) is used. In Theorem 3.2.1 below, we establish convergence in law to the invariant measure for the Langevin dynamics (3.1.1).
We will pose the following assumptions on U:

Assumption 5. The function $U \in C^\infty(\mathbb{R}^n)$ satisfies $U \geq 0$. Moreover, there exist constants $\beta_1, \beta_2 > 0$ and $\alpha \in \mathbb{R}$ such that

$$\forall q \in \mathbb{R}^n, \langle q, \nabla_q U(q)\rangle \geq \beta_1 U(q) + \beta_2 |q|^2 + \alpha. \tag{3.2.3}$$

The following Lyapunov function $\mathcal{K}_l : \mathbb{R}^{2n} \to \mathbb{R}$ for all $l \in \mathbb{N}$ will be used:

$$\mathcal{K}_l(z) = \mathcal{K}_l(q,p) = \Big(cU(q) + a|q|^2 + b\langle q, p\rangle + \frac{c}{2}|p|^2 + 1\Big)^l \tag{3.2.4}$$

for constants $a, b, c > 0$. The well-posedness of equation (3.1.1) is stated in Theorem 3.6.1.

Theorem 3.2.1. *Under Assumption 5, $\tilde{\pi}$ is the unique invariant probability measure for the SDE* (3.1.1) *and for all $l \in \mathbb{N}$, there exist constants $\kappa_l, C_l > 0$ depending on l and constants $a, b, c > 0$ independent of l such that the solution $z_t^z = (q_t, p_t)$ to* (3.1.1) *with initial condition z satisfies*

$$|\mathbb{E}[\varphi(z_t^z)] - \tilde{\pi}(\varphi)| \leq C_l e^{-t\kappa_l} \mathcal{K}_l(z) \left\| \frac{\varphi - \tilde{\pi}(\varphi)}{\mathcal{K}_l} \right\|_{L^\infty} \tag{3.2.5}$$

for Lebesgue almost all initial $z \in \mathbb{R}^{2n}$, $\mathcal{K}_l \geq 1$ given by (3.2.4) *and all Lebesgue measurable φ satisfying*

$$\frac{\varphi}{\mathcal{K}_l} \in L^\infty \tag{3.2.6}$$

Moreover for any $l \in \mathbb{N}$, $\mathcal{K}_l$ satisfies

$$\int \mathcal{K}_l d\tilde{\pi} < \infty \tag{3.2.7}$$

and

$$\mathcal{L}\mathcal{K}_l \leq -a_l \mathcal{K}_l + b_l \tag{3.2.8}$$

for some constants $a_l, b_l > 0$.

For the sake of brevity we omit the proof. The fact that $\tilde{\pi}$ is invariant is thanks to (3.2.2). For the rest of the statements, the proof is contained in [172, Theorem 3], which is based on [133]. In the latter the setting is more general than (3.1.1) in that the friction matrix is dependent on q and the drift is not necessarily conservative, i.e. the forcing term is not the gradient of a scalar function and the fluctuation-dissipation theorem (see equation (6.2) in [158]) does not hold, but of course, [172, Theorem 3] applies in particular to our setting.

Remark 3.2.1. Inequality (3.2.5) holds for all initial $z \in \mathbb{R}^{2n}$, as opposed to almost all z, given any *bounded* measurable φ. This is a consequence of combining (3.2.5) together with the strong Feller property given by Theorem 4.2 in [51].

The following corollary holds by taking φ as indicator functions and Remark 3.2.1.

Corollary 3.2.2. *Under Assumption 5, for all initial $z \in \mathbb{R}^{2n}$, the transition probability ρ_t^z of* (3.1.1)*, given by $\rho_t^z(A) = \mathbb{P}(z_t^z \in A)$, satisfies*

$$\|\rho_t^z - \tilde{\pi}\|_{TV} \to 0 \qquad \text{as } t \to \infty$$

where $\|\cdot\|_{TV}$ denotes the total variation norm.

The solution to the Poisson equation is given next following the direction of [29], see the paragraph before Corollary 3.2 there.

Theorem 3.2.3. *Under Assumption 5, if $f \in L_0^2(\tilde{\pi})$ satisfies $\frac{f}{\mathcal{K}_l} \in L^\infty$ for some $l \in \mathbb{N}$, then there exists a unique solution $\phi \in L_0^2(\tilde{\pi})$ to the Poisson equation* (3.1.5)*. Moreover, the solution is given by*

$$\phi = \int_0^\infty P_t(f)dt. \tag{3.2.9}$$

Proof. For $T > 0$, let $g_T := \int_0^T P_t(f)dt$. Note that $g_T \in L^2(\tilde{\pi})$ for $T \in \mathbb{R}_+ \cup \{\infty\}$ and by Theorem 3.2.1

$$g_T \to \int_0^\infty P_t(f)dt \tag{3.2.10}$$

in $L^2(\tilde{\pi})$ as $T \to \infty$, specifically (3.2.5) with $\varphi = f$ and using (3.2.7) for $2l$ in place of l. Applying L, it holds that

$$Lg_T = \lim_{s\to 0} \frac{P_s(g_T) - g_T}{s} = \lim_{s\to 0} \frac{1}{s}\left(\int_s^{T+s} - \int_0^T\right) P_u(f)du = P_T(f) - f,$$

where the exchange in the order of integration is justified by Fubini, (3.2.5) and the last equality follows by the strong continuity of $(P_t)_{t\geq 0}$ (given by Proposition 3.6.2). Inequalities (3.2.5) and (3.2.7) (with $2l$ in place of l) also give

$$P_T(f) \to 0 \qquad \text{in } L^2(\tilde{\pi}) \tag{3.2.11}$$

as $T \to \infty$, so that since L is a closed operator, equations (3.1.5) and (3.2.9) hold. In addition, $\int \phi d\tilde{\pi} = 0$ follows from the invariance of $\tilde{\pi}$, Theorem 3.2.1 and Fubini's theorem. □

We proceed to state the central limit theorem for the solution to (3.1.1).

Theorem 3.2.4. *Under Assumption 5, if $f \in L^2(\tilde{\pi})$ satisfies $\frac{f}{\mathcal{K}_l} \in L^\infty$ for some $l \in \mathbb{N}$, the random variable $\frac{1}{\sqrt{t}}\int_0^t (f(z_s) - \pi(f))ds$ converges in distribution to $\mathcal{N}(0, \sigma_f^2)$ as $t \to \infty$ for any initial distribution, where*

$$\sigma_f^2 = 2\int \phi(f - \pi(f))d\tilde{\pi} \tag{3.2.12}$$

and $\phi \in L_0^2(\tilde{\pi})$ is the solution to (3.1.5)*.*

Proof. By Corollary 3.2.2 and Theorem 3.2.3, the result follows by Theorem 2.6 in [15]. Note that the joint measurability assumption in [15] of the transition probability is verified in Theorem 3.6.1. See also Theorem 3.1 in [29]. □

3.3 Directional derivative of σ^2

In this section, we give a number of natural preliminary results that pave the path for the main result in Theorem 3.3.2, in which a formula for the derivative (3.1.6) of σ^2 with respect to Γ is provided. The derivative formula is based on the fact that the invariant measure $\tilde{\pi}$ is independent of Γ appearing in the Langevin equation (3.1.1), yet on the other hand that the asymptotic variance varies as Γ changes. It gives a systematic method to find an optimal parameter in our dynamics, in contrast to the numerical comparison in the previous Chapter 2 (see Figure 2.3.2). The proofs of Lemma 3.3.1 and Theorem 3.3.2 are deferred to Section 3.6.

3.3.1 Preliminary results and the main formula

In order for the integral in a formula like (3.1.6) to be finite, control on the derivatives in p is required. This will also be used in the proof of Theorem 3.3.2 and it is given by the following Lemma 3.3.1. Note that the results in the seminal work of Pardoux and Veretennikov are not relevant for our derivative bounds for a number of reasons, for example results in [154] may not be applied because the diffusion matrix in (3.1.1) is degenerate.

Lemma 3.3.1. *Under Assumption 5, if $f \in L_0^2(\tilde{\pi})$ satisfies $\frac{f}{\mathcal{K}_l} \in L^\infty$ for some $l \in \mathbb{N}$, then the weak derivative in p of the solution ϕ to $-L\phi = f$, denoted $\nabla_p \phi$, satisfies $\int |\nabla_p \phi|^2 d\tilde{\pi} < \infty$.*

The main result of this section is the expression for the directional derivative of the asymptotic variance and is given next. Since Lemma 3.3.1 is available only for observable functions of position q, the formula for the derivative is given for such observables. The directional derivative of $E : \mathbb{S}^n_{++} \to \mathbb{R}$ at $\Gamma \in \mathbb{S}^n_{++}$ in a symmetric matrix direction $\delta\Gamma \in \mathbb{R}^{n\times n}$ is denoted by $dE(\Gamma).\delta\Gamma = \lim_{\epsilon\to 0} \frac{1}{\epsilon}(E(\Gamma + \epsilon\delta\Gamma) - E(\Gamma))$ whenever the limit exists. The explicit dependence on Γ is omitted in the notation when it is clear from the context.

Theorem 3.3.2. *Under Assumption 5, if $f = f(q) \in L_0^2(\pi)$ is continuous, satisfies $\frac{f}{\mathcal{K}_l} \in L^\infty$ for some $l \in \mathbb{N}$ and there exists $\epsilon' > 0$ such that $\Gamma, \Gamma + \epsilon\delta\Gamma \in \mathbb{S}^n_{++}$ for $|\epsilon| \le \epsilon'$, then the directional derivative of the asymptotic variance σ^2 at Γ in the direction $\delta\Gamma \in \mathbb{R}^{n\times n}$ has the form*

$$d\sigma^2(\Gamma).\delta\Gamma = -2\int (\nabla_p \phi)^\top \delta\Gamma \nabla_p \tilde{\phi} d\tilde{\pi}, \tag{3.3.1}$$

where ϕ is the solution (3.2.9) *to the Poisson equation for the dynamics* (3.1.1) *at Γ and $\tilde{\phi}$ is given by* (3.1.7)*.*

As mentioned in the introduction, from (3.3.1), the direction (3.1.8) guarantees a decrease in asymptotic variance; similarly the scalar change in Γ given by (3.1.8) where the outer product is replaced by a dot product guarantees a decrease in σ^2.

3.3.2 A formula using a tangent process

The directional derivative $d\sigma^2.\delta\Gamma$ of the asymptotic variance (3.3.1) can be written in a more useful form for simulation based approximation. The first variation process of (3.1.1) is used here to calculate (3.1.10); this will be the main methodology used in the numerical sections. This alternative formula given in Theorem 3.3.3 provides a way to avoid using a finite difference Monte Carlo estimate of the derivative of an expectation. For simplicity, we set $M = I_n$ here. The first variation process (with respect to initial momenta p) associated to (3.1.1), denoted by $(D_pq_t, D_pp_t) \in \mathbb{R}^{n\times 2n}$ for $t \geq 0$, is defined as the matrix-valued solution to

$$\partial_t \begin{pmatrix} D_pq_t \\ D_pp_t \end{pmatrix} = \begin{pmatrix} 0 & I_n \\ -D^2U(q_t) & -\Gamma \end{pmatrix} \begin{pmatrix} D_pq_t \\ D_pp_t \end{pmatrix} \tag{3.3.2}$$

with the initial condition $D_pq_0 = 0$, $D_pp_0 = I_n$. By Theorems V.39 in [165], the partial derivatives of (q_t, p_t) with respect to the initial values in p is the unique solution to (3.3.2) and (D_pq_t, D_pp_t) is continuous with respect to those initial values. The aim in setting the various assumptions in the following Theorem 3.3.3 is to ensure the exponential decay of these derivative processes, as opposed to exponential decay of derivatives of the associated semigroup as in [45]. We omit in the notational dependence of (q_t, p_t) on its initial condition $(q_0, p_0) = (q, p) = z$ whenever convenient in the following.

Theorem 3.3.3. *Let Assumption 5 hold. If in addition,*

- *there exist $U_0 > 0$ and $Q \in \mathbb{S}^n_{++}$ such that for all $q \in \mathbb{R}^n$, $v \in \mathbb{R}^n$,*

$$v^\top D^2U(q)v \geq U_0|v|^2, \quad D^2U(q) = Q + D(q),$$

  where $D : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is small enough everywhere in the following sense:

$$|D(q)| \leq \hat{\lambda} := \min\left(\frac{\lambda_m}{2}, \frac{\lambda_m U_0^2}{8\lambda_M^2}, \frac{\lambda_m U_0}{16}, \frac{U_0}{8}\sqrt{\sigma_{\min}(Q)}\right), \tag{3.3.3}$$

  where $\lambda_m, \lambda_M > 0$ are respectively the smallest and largest eigenvalue of Γ and $\sigma_{\min}(Q)$ denotes the smallest eigenvalue of Q;

- *$f = f(q) \in L_0^2(\pi) \cap C^1(\mathbb{R}^n)$ and satisfies $\frac{|f|+|\nabla f|}{\mathcal{K}_l} \in L^\infty$ for some $l \in \mathbb{N}$,* *then the weak derivative* $\nabla_p \phi$ *has the form*

$$\nabla_p \phi(q,p) = \int_0^\infty \mathbb{E}[\nabla f(q_t)^\top D_p q_t] dt, \tag{3.3.4}$$

where q_t *solves* (3.1.1) *with initial condition* $(q_0, p_0) = (q, p)$ *and* $D_p q_t$ *solves* (3.3.2)*, the latter satisfying*

$$|D_p q_t|^2 + |D_p p_t|^2 \leq C' e^{-Ct} \tag{3.3.5}$$

for some constants $C, C' > 0$ *independent of* (q_0, p_0) *and* $\omega \in \Omega$*.*

The additional assumptions on U are made in order to ensure that the process $(D_p q_t, D_p p_t)$ converges to zero exponentially quickly so that the integral in (3.3.4) is finite. In particular, (3.3.3) requires U to be close to a quadratic function $q^\top Q q$; see also [21] for a situation where a similar assumption is made for the long time behaviour for the Vlasov-Fokker-Planck equation.

Remark 3.3.1. Exponential decay of the first variation process is not required for the integrability of the integrand on the right-hand side of (3.3.4), only some uniform (in initial condition (q_0, p_0)) integrability in time of $D_p q_t$ together with a boundedness assumption on ∇f. On the other hand, Proposition 1 in [47] and Proposition 4 in [142] explores more detailed conditions under which contractivity holds and does not hold for scalar friction $\Gamma = \gamma > 0$; our result places the focus on conditions on U such that contractivity holds for all $\Gamma \in S^n_{++}$.

Proof. Let $b > 0$ be the constant

$$b = \min\left(\frac{\lambda_m U_0}{2\lambda_M^2}, \frac{\lambda_m}{4}, \frac{1}{2}\sqrt{\sigma_{\min}(Q)}\right) \tag{3.3.6}$$

so that $\hat{\lambda}$ reduces to $\hat{\lambda} = \min\left(\frac{\lambda_m}{2}, b\frac{U_0}{4}\right)$ and, since $b \leq \frac{1}{2}\sqrt{\sigma_{\min}(Q)}$, the matrix $\begin{pmatrix} Q & bI_n \\ bI_n & I_n \end{pmatrix}$

is positive definite. We have the following bound.

$$
\begin{aligned}
&\frac{1}{2}\partial_t\left[e_i^\top \begin{pmatrix} D_p q_t \\ D_p p_t \end{pmatrix}^\top \begin{pmatrix} Q & bI_n \\ bI_n & I_n \end{pmatrix} \begin{pmatrix} D_p q_t \\ D_p p_t \end{pmatrix} e_i\right] \\
&= e_i^\top D_p q_t^\top Q D_p p_t e_i + b|D_p p_t e_i|^2 \\
&\quad - e_i^\top (bD_p q_t + D_p p_t)^\top (D^2 U(q) D_p q_t + \Gamma D_p p_t) e_i \\
&= -b e_i^\top D_p q_t^\top D^2 U(q_t) D_p q_t e_i + e_i^\top D_p q_t^\top (-b\Gamma - D(q_t)) D_p p_t e_i \\
&\quad - e_i^\top D_p p_t^\top (\Gamma - bI_n) D_p p_t e_i \\
&\leq \left(-bU_0 + \frac{bU_0}{2} + \frac{\hat{\lambda}}{2}\right)|D_p q_t e_i|^2 + \left(-\lambda_m + b + \frac{b\lambda_M^2}{2U_0} + \frac{\hat{\lambda}}{2}\right)|D_p p_t e_i|^2 \\
&\leq -\frac{bU_0}{4}|D_p q_t e_i|^2 - \frac{\lambda_m}{4}|D_p p_t e_i|^2 \\
&\leq -C e_i^\top \begin{pmatrix} D_p q_t \\ D_p p_t \end{pmatrix}^\top \begin{pmatrix} Q & bI_n \\ bI_n & I_n \end{pmatrix} \begin{pmatrix} D_p q_t \\ D_p p_t \end{pmatrix} e_i
\end{aligned} \tag{3.3.7}
$$

for some generic constant $C > 0$ independent of the initial values (q_0, p_0) and $\omega \in \Omega$. Consequently, using the weighted boundedness assumption on $|\nabla f|$ (that $\frac{|\nabla f|}{\mathcal{K}_l} \in L^\infty$ for some $l \in \mathcal{N}$) and for each index i,

$$
\begin{aligned}
\left|(\nabla f(q_t)^\top D_p q_t)_i\right| &\leq C' e^{-Ct}|\nabla f(q_t)| \\
&\leq C' e^{-Ct}(|\nabla f(q_t)| - \pi(|\nabla f|)) + C' e^{-Ct}
\end{aligned} \tag{3.3.8}
$$

for a generic $C' > 0$ independent of (q_0, p_0) and $\omega \in \Omega$. Due to (3.3.8) together with Fubini's theorem, it holds for $T > 0$ and a test function $g \in C_c^\infty(\mathbb{R}^{2n})$ that

$$
\begin{aligned}
\int\int_0^T \mathbb{E}[f(q_t^z)]dt\nabla_p g(z)dz &= \int_0^T \mathbb{E}\left[\int f(q_t^z)\nabla_p g(z)dz\right]dt \\
&= -\int_0^T \mathbb{E}\left[\int \nabla f(q_t^z)^\top D_p q_t^z g(z)dz\right]dt \\
&= -\int\int_0^T \mathbb{E}[\nabla f(q_t^z)^\top D_p q_t^z]dt g(z)dz.
\end{aligned}
$$

Using Theorem 3.2.1, (3.3.8) again and dominated convergence to take $T \to \infty$ on both sides concludes the proof. □

The following is a brief discussion about how equation (3.3.4) can be used in practice to approximate the gradient direction $\int \nabla_p \phi \otimes \nabla_p \tilde{\phi} d\tilde{\pi}$ from realisations (q_t, p_t) of (3.1.1). We have in mind at first setting $(q^*, p^*) = (q_0, p_0)$, where (q_0, p_0) is the initial condition from equation (3.1.1), so that equation (3.3.4) implies $\nabla_p \phi(q^*, p^*) \otimes \nabla_p \tilde{\phi}(q^*, p^*)$ can be approximated using

$$\delta\Gamma = \int_0^T \nabla f(q_s^{(q^*,p^*)})^\top D_p q_s^{(q^*,p^*)} ds \otimes \int_0^T \nabla f(q_s^{(q^*,-p^*)})^\top D_p q_s^{(q^*,-p^*)} ds, \tag{3.3.9}$$

where $(q_s^{(q^*,p^*)}, p_s^{(q^*,p^*)})$ solves (3.1.1) with initial condition (q^*, p^*), and $(q_s^{(q^*,-p^*)}, p_s^{(q^*,-p^*)})$ denotes a parallel solution of (3.1.1) with initial condition $(q^*, -p^*)$ and independent realisations for W. Moreover, $(D_p q_s)_{0\leq s\leq T}$ denotes corresponding solutions to (3.3.2) for both initial conditions. At time T one can then update Γ using equation (3.3.9), update $(q^*, p^*) = (q_T, p_T)$ and repeat using (3.3.9) in the same way until some satisfactory Γ has been reached. There are sources of bias from not being at stationarity and using finite T, but both these can be mitigated in practice with careful and adaptive choice of T. The overall approach is summarised in Algorithm 1 and given with more detail in Algorithm 2. The next result is that the estimator (3.3.9) has finite variance.

Theorem 3.3.4. *Let the assumptions of Theorem 3.3.3 hold. For Lebesgue almost-all $(q,p) \in \mathbb{R}^{2n}$, each entry of $\delta\Gamma$ defined in* (3.3.9) *has finite variance.*

Proof. It suffices to show that (3.3.9) has finite second moment, for which it suffices to show that each element in the vector of time integrals $\int_0^T \nabla f(q_t^{(q,p)})^\top D_p q_t^{(q,p)} ds$ has finite second moments by independence. For each index i, using (3.3.7),

$$\begin{aligned}\left|(\nabla f(q_t)^\top D_p q_t)_i\right|^2 &\leq C'^2 e^{-2Ct} |\nabla f(q_t)|^2 \\ &\leq C'^2 e^{-2Ct} (|\nabla f(q_t)|^2 - \pi(|\nabla f|^2)) + C'^2 e^{-2Ct} \pi(|\nabla f|^2),\end{aligned}$$

so that using the (weighted) boundedness assumption on $|\nabla f|$ together with Theorem 3.2.1 and Fubini's theorem, the proof concludes. □

3.4 Gaussian cases

Throughout this section, the target measure π is assumed to be Gaussian, when π is mean zero this is $\pi \propto \exp(-\frac{1}{2} q^\top \Sigma^{-1} q)$ for $\Sigma \in \mathbb{S}^n_{++}$, in other words, the potential is quadratic, $U(q) = \frac{1}{2} q^\top \Sigma^{-1} q$. For polynomial function observables, we look for solutions to the Poisson equation (3.1.5) by using a polynomial ansatz and comparing coefficients in order to obtain an explicit expression for the asymptotic variance. The results provide benchmarks to test the performance of the algorithms that arise from using the gradient in Theorem 3.3.2 as well as intuition for how Γ can be improved in concrete cases. We will consider the following cases for the observables:

1. Quadratic $f = \frac{1}{2} q^\top U_0 q$ under the assumption of commutativity between U_0 and Σ (Proposition 3.4.5), also $f = \frac{1}{2} U_0 q^2 + lq$ in one dimension (Proposition 3.4.6);

2. Odd polynomial f, where the asymptotic variance will be shown to decrease to zero as $\Gamma \to 0$ (Proposition 3.4.7, Corollary 3.4.8 and Proposition 3.4.9);

3. Quartic f in one dimension, in which case the situation is similar to quadratic f (Proposition 3.4.10).

4. Quadratic f, but with polynomial Γ. This is not the setting above, but the difficulty of any such extension to the case of variable friction Γ is demonstrated by a negative result about the form of optimal Γ in polynomial settings, in which one hopes for analytical solutions to the Poisson equation.

All of the proofs and derivations for the results in this section can be found in Section 3.6. We proceed with stating in more detail the general situation of this section.
Let $\Sigma \in \mathbb{S}^n_{++}$, $U_0 \in \mathbb{S}^n_{++}$ and $l \in \mathbb{R}^n$. The Gaussian invariant measure $\tilde{\pi}$ and the observable $f : \mathbb{R}^{2n} \to \mathbb{R}$ are given by

$$\tilde{\pi} \propto \exp\left(-\frac{1}{2}q \cdot \Sigma^{-1} q - \frac{1}{2} p \cdot M^{-1} p\right), \quad f(q) = \frac{1}{2} q \cdot U_0 q + l \cdot q \tag{3.4.1}$$

and the value $\pi(f)$ becomes

$$\pi(f) = \int f d\pi = \int \frac{1}{2} q \cdot U_0 q d\pi = \frac{1}{2} U_0 : \Sigma. \tag{3.4.2}$$

The infinitesimal generator $\mathcal{L}$ becomes in this case

$$\begin{aligned}
\mathcal{L} &= \begin{pmatrix} 0 & M^{-1} \\ -\Sigma^{-1} & -\Gamma M^{-1} \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \cdot \nabla + \nabla_p \cdot \Gamma \nabla_p \\
&= M^{-1} p \cdot \nabla_q - \Sigma^{-1} q \cdot \nabla_p - \Gamma M^{-1} p \cdot \nabla_p + \nabla_p \cdot \Gamma \nabla_p.
\end{aligned} \tag{3.4.3}$$

Consider the natural candidate solution ϕ to the Poisson equation (3.1.5) given by

$$\phi(q,p) = \frac{1}{2} q \cdot G q + q \cdot E p + \frac{1}{2} p \cdot H p + g \cdot q + h \cdot p - \frac{1}{2}(G : \Sigma + H : M). \tag{3.4.4}$$

for some constant matrices $G, E, H \in \mathbb{R}^{n \times n}$ and vectors $g, h \in \mathbb{R}^n$. Note that we allow G and H not to be symmetric and specify G_S and H_S as the respective symmetric parts in order to make a clear distinction.

Lemma 3.4.1. *Given f in* (3.4.1), *$\pi(f)$ in* (3.4.2) *and $\mathcal{L}$ of the form* (3.4.3), *it holds* *that ϕ given by* (3.4.4) *is a solution to the Poisson equation* (3.1.5) *if and only if*

$$\Sigma^{-1}q\cdot(E^\top q+h)-\Gamma:H_S-\frac{1}{2}q\cdot U_0q-l\cdot q+\frac{1}{2}U_0:\Sigma=0, \tag{3.4.5}$$

$$-M^{-1}(G_Sq+g)+H_S\Sigma^{-1}q+M^{-1}\Gamma(E^\top q+h)=0, \tag{3.4.6}$$

$$-E^\top M^{-1}+H_S\Gamma M^{-1}=A_1, \tag{3.4.7}$$

for some antisymmetric $A_1\in\mathbb{R}^{n\times n}$.

3.4.1 Quadratic observable

Similar calculations in this situation have appeared previously in Proposition 1 in [55], where explicit expressions analogous to G, E, H and for σ^2 are given. For our purposes of finding an optimal Γ, the approach take here is different. Instead of taking these explicit expressions, we keep unknown antisymmetric matrices (such as A_1) as they appear as an alternative to the aforementioned explicit expressions. Eventually the commutativity property between Σ and U_0 is used to show that the antisymmetric matrices are zero. We continue from (3.4.5), (3.4.6) and (3.4.7) with finding explicit expressions for the coefficients G, E, H of ϕ.

Lemma 3.4.2. *Given f in* (3.4.1)*, $\pi(f)$ in* (3.4.2)*, $\mathcal{L}$ of the form* (3.4.3)*, ϕ given by* (3.4.4) *is a solution to the Poisson equation* (3.1.5) *with* (3.4.3) *if and only if there exist antisymmetric matrices A_1, A_2 such that*

$$G_S=\frac{1}{2}M(\Sigma U_0-\Sigma A_2-2A_1M)\Gamma^{-1}\Sigma^{-1}+\frac{1}{2}\Gamma(U_0\Sigma-A_2\Sigma), \tag{3.4.8}$$

$$E=\frac{1}{2}U_0\Sigma+\frac{1}{2}A_2\Sigma, \tag{3.4.9}$$

$$H_S=\frac{1}{2}(\Sigma U_0-\Sigma A_2-2A_1M)\Gamma^{-1}, \tag{3.4.10}$$

$$h=\Sigma l \qquad \textit{and} \qquad g=\Gamma\Sigma l. \tag{3.4.11}$$

The asymptotic variance from Theorem 3.2.4 can be given by a formula in terms of Σ, U_0 and the coefficients of ϕ. Before substituting the expressions from Lemma 3.4.2 into the formula, we give the formula itself, which is adapted from the proof of Proposition 1 in [55].

Lemma 3.4.3. *Let f be given by* (3.4.1)*, $\pi(f)$ be given by* (3.4.2) *and $\mathcal{L}$ be given by* (3.4.3)*. If the solution ϕ to the Poisson equation* (3.1.5) *is of the form* (3.4.4)*, then the asymptotic variance σ^2 given by* (3.2.12) *has the expression*

$$2\langle\phi,f-\pi(f)\rangle_{\tilde{\pi}}=\mathit{Tr}(G_S\Sigma U_0\Sigma)+2g\cdot\Sigma l. \tag{3.4.12}$$

From the expressions (3.4.8) and (3.4.10) for G_S and H_S respectively, it is not straightforward to check that there exist antisymmetric A_1 and A_2 such that the right hand sides are indeed symmetric at this point, which is necessary for the ansatz (3.4.4) for ϕ to be a valid solution. On the other hand, if Σ, U_0, Γ, M all commute, then the right hand sides of (3.4.8) and (3.4.10) are symmetric for $A_1 = A_2 = 0$ and the coefficients G and H become explicit, which allows taking derivatives of σ^2 with respect to the entries of Γ. Moreover, the explicit coefficients allow optimisation of M, which is given by the following proposition.

Proposition 3.4.4. *Suppose Σ, U_0 and Γ all commute. Let f be as in* (3.4.1)*, $\pi(f)$ be as in* (3.4.2)*, $\mathcal{L}$ be of the form* (3.4.3) *and ϕ be the solution to the Poisson equation* (3.1.5)*. It holds that*

$$\lim_{M=mI_n, m\downarrow 0} \int \phi(f-\pi(f))d\tilde{\pi} = \inf_{M\in\mathbb{S}_\Sigma} \int \phi(f-\pi(f))d\tilde{\pi}, \tag{3.4.13}$$

where $\mathbb{S}_\Sigma$ is the set of symmetric positive definite matrices commuting with Σ.

Remark 3.4.1. The limit (3.4.13) in Proposition 3.4.4 is, together with a rescaling in the velocity space, the overdamped limit of the Langevin dynamics, see Section 2.2.4 in [123]. However, (3.4.13) does not necessarily mean overdamped dynamics are better in practice. For example when Γ is a small scalar, the overdamped limit corresponding to (3.4.13) results in a time speed-up inversely proportional to Γ over the overdamped limit corresponding to $\Gamma = I_n$. Consequently, any such comparison between Langevin dynamics and the overdamped limit should include constraints such as those in [81] for both sets of dynamics. We focus on the optimisation of Γ and fix $M = I_n$ in the following.

As before, we denote $\mathbb{S}_\Sigma$ to be the set of symmetric positive definite matrices commuting with Σ.

Proposition 3.4.5. *Let Σ, U_0, l, M be such that*

$$\Sigma U_0 = U_0\Sigma, \qquad l = 0, \qquad M = I_n, \tag{3.4.14}$$

the function f be as in (3.4.1)*, $\pi(f)$ be as in* (3.4.2)*, $\mathcal{L}$ be of the form* (3.4.3) *and ϕ be the solution to the Poisson equation* (3.1.5)*. It holds that*

$$\min_{\Gamma\in\mathbb{S}_\Sigma} 2\int \phi(f-\pi(f))d\tilde{\pi} = Tr\Big(U_0^2\Sigma^{\frac{5}{2}}\Big),$$

where the minimum is attained by $\Gamma = \Sigma^{-\frac{1}{2}}$.

In the scalar case, we can remove the restriction on l.

Proposition 3.4.6. *If $n = 1$, $U_0 \neq 0$, $l \neq 0$, $f : \mathbb{R} \to \mathbb{R}$ is given by* (3.4.1)*, $\pi(f)$ is given by* (3.4.2)*, $\mathcal{L}$ is of the form* (3.4.3) *and ϕ is the solution to the Poisson equation* (3.1.5)*, then $\min_{\Gamma>0} 2\int \phi(f - \pi(f))d\tilde{\pi} = M^{\frac{1}{2}}\Sigma^2 U_0^2(\Sigma + 4l^2 U_0^{-2})^{\frac{1}{2}}$ and the minimum is attained by $\Gamma = \frac{M^{\frac{1}{2}}}{(\Sigma+4l^2U_0^{-2})^{\frac{1}{2}}}$.*

3.4.2 Odd polynomial observable

Another special case within (3.4.1) where the solution ϕ can be readily identified is when $U_0 = 0$, that is, for linear observables. More generally, (almost) zero variance can be attained in the following special case.

Proposition 3.4.7. *Under Assumption 5, for a general target measure $\pi \propto e^{-U}$ on $\mathbb{R}^n$, if the observable f is of the form $f(q) = \alpha \cdot \nabla U$, for $\alpha = (\alpha_1, \dots, \alpha_n)$, $\alpha_i \in \mathbb{R}$, $\mathcal{L}$ is of the general form* (3.2.1) *and ϕ is the solution to the Poisson equation* (3.1.5)*, then the asymptotic variance satisfies*

$$\inf_{\Gamma \in \{\gamma I_n : \gamma > 0\}} 2 \int \phi(f - \pi(f))d\tilde{\pi} = 0. \tag{3.4.15}$$

Corollary 3.4.8. *Given a Gaussian target measure with density $\pi \propto e^{-U}$ on $\mathbb{R}^n$, observable $f : \mathbb{R}^n \to \mathbb{R}$ as in* (3.4.1) *with $U_0 = 0$, that is, $f(q) = l \cdot q$, where $l \in \mathbb{R}^n$, $\pi(f) = 0$, $\mathcal{L}$ of the form* (3.2.1) *and ϕ the solution to the Poisson equation* (3.1.5)*, equation* (3.4.15) *holds.*

Note that Corollary 3.4.8 is also a consequence of (3.6.19) in the proof of Lemma 3.4.3. Furthermore, the setting in Corollary 3.4.8 is included in Proposition 3.4.4, which suggests that the overdamped limit (see Remark 3.4.1) for some fixed small Γ (used to obtain arbitrarily small asymptotic variance in the proof of Proposition 3.4.7) also has small asymptotic variance. One can readily check, at least formally, that the smallness of Γ corresponds to a speed up in time for the overdamped limit.

We give here some intuition for the situation in Corollary 3.4.8. First note that the Langevin diffusion with $\Gamma = 0$ reduces to deterministic Hamiltonian dynamics and that it is the limit case for the Γ attaining arbitrarily small asymptotic variance in the proof of Proposition 3.4.7. The result indicates that this is optimal in the linear observable, Gaussian measure case (i.e. (3.4.1), $U_0 = 0$) and this aligns with the fact that the value (3.4.2) to be approximated is exactly the value at the $q = p = 0$, so that Hamiltonian dynamics starting at $q = 0$, staying there for all time, approximates the integral (3.4.2) with perfect accuracy. A similar idea holds for when the initial condition is not $q = p = 0$, where (3.4.2) is approximated exactly after any integer number of orbits in (q, p) space. Continuing on this idea, it seems reasonable that the same statement holds more generally for any odd observable. At least, the following holds in one dimension.

Proposition 3.4.9. *If $n = 1$, $\hat{k} \in \mathbb{N}_0$ and $f : \mathbb{R} \to \mathbb{R}$ is an odd finite order polynomial observable given by*

$$f(q) = \sum_{i=0}^{\hat{k}} a_i q^{2i+1}, \tag{3.4.16}$$

also $\pi(f) = 0$, $\mathcal{L}$ is of the form (3.4.3) *and ϕ is the solution to the Poisson equation* (3.1.5)*, then the asymptotic variance satisfies* (3.4.15)*.*

3.4.3 Quartic observable

The situation in the quartic observable case, at least in one dimension, is similar to quadratic observable case.

Proposition 3.4.10. *If $n = 1$ and $f : \mathbb{R} \to \mathbb{R}$ is a quartic observable given by*

$$f(q) = q^4, \tag{3.4.17}$$

also $\pi(f) = 3\Sigma^2$ for some $\Sigma > 0$, $\mathcal{L}$ is of the form (3.4.3)*, $M = 1$ and ϕ is the solution to the Poisson equation* (3.1.5)*, then there exists $\sigma_{quar} > 0$ such that* $\min_{\Gamma=\gamma>0} 2\int \phi(f - \pi(f))d\tilde{\pi} = \sigma_{quar}$*.*

3.4.4 Polynomial Γ

The final consideration in this section is variable friction $\Gamma : \mathbb{R}^n \to \mathbb{R}$ in the case of (3.4.1). Although Langevin dynamics with variable friction is not the setting of this chapter, a natural class of functions for Γ where the Poisson equation might be expected to have a closed form solution is when Γ is polynomial; we demonstrate the difficulty of this setting by presenting here the negative result that no finite order polynomial ϕ solves the Poisson equation for $n = 1$.

Proposition 3.4.11. *If $n = 1$, $f : \mathbb{R} \to \mathbb{R}$ is given by* (3.4.1) *with $U_0 \neq 0$, $\pi(f)$ is given by* (3.4.2)*, $\mathcal{L}$ is given by* (3.4.3)*, Γ is given by a nonconstant finite order polynomial, then finite order polynomials in q, p cannot be a solution ϕ to the Poisson equation* (3.1.5)*.*

The proof of Proposition 3.4.11 can be found in Section 3.6.

3.5 Computation of the change in Γ

Throughout this section, the $M = I_n$ case is considered. As mentioned, the formula (3.3.1) gives a natural gradient descent direction (3.1.8) to take Γ in order to optimise σ^2 from Theorem 3.2.4. In Theorem 3.3.2 and in the form (3.1.8), the expression for the gradient is already susceptible to a Green-Kubo approach in the sense that the form (3.2.9) for ϕ

can be substituted in to obtain a trajectory based formula, where finite difference is used to approximate ∇_p and independent realisations of (q_t, p_t) is used for the expectations. However, this is too inaccurate in the implementation to be useful. The more directly calculable form as stated in the introduction in (3.1.10) is used involving the derivative of (q_t, p_t) with respect to the initial condition in Section 3.3.2.

We focus the discussion on a Monte Carlo method to approximate $\nabla\phi$ and gradient directions in Γ (e.g. (3.1.8)) based on Theorem 3.3.2, but a spectral method to solve (3.1.5) and compute the change in Γ is given in Section 3.5.5, which is computationally feasible in low dimensions. Algorithm 1 summarises the resulting procedure, where all expectations within (3.1.8) are approximated by single realisations; further justifications, refinements and a concrete implementation (Algorithm 2) can be found in Section 3.5.1, whilst alternative methods are given in Section 3.5.5.

Algorithm 1 Continuous-time outline of Γ update using (3.1.6) and (3.1.10)

Result: $\Gamma \in \mathbb{S}^n_{++}$
Start from arbitrary $(q_0, p_0) \in \mathbb{R}^{2n}$ and set $(\tilde{q}_0, \tilde{p}_0) = (q^0, -p^0)$, $Dq_0 = D\tilde{q}_0 = 0$, $Dp_0 = D\tilde{p}_0 = I_n$, $\zeta = \tilde{\zeta} = 0$, $\Gamma = I_n$, $t = t_0 = 0$;
for N epochs **do**
  simulate one time-step in q_t, $\tilde{q}_t$ then in $D_p q_t$ and $D_p \tilde{q}_t$;
  add to $\zeta, \tilde{\zeta}$ to approximate the row vectors

$$\zeta = \int_{t_0}^{t} \nabla f(q_s)^\top D_p q_s ds, \quad \tilde{\zeta} = \int_{t_0}^{t} \nabla f(\tilde{q}_s)^\top D_p \tilde{q}_s ds;$$

  if $(D_p q_t, D_p p_t)$ is small enough in magnitude **then**
    update Γ with the gradient direction $-\zeta \otimes \tilde{\zeta} - (\zeta \otimes \tilde{\zeta})^\top$;
    reset $(\tilde{q}_t, \tilde{p}_t) \leftarrow (q_t, -p_t)$; $(D_p q_t, D_p p_t), (D_p \tilde{q}_t, D_p \tilde{p}_t) \leftarrow (0, I_n)$; $t_0 \leftarrow t$; $\zeta, \tilde{\zeta} \leftarrow 0$;
  end if
  $t \leftarrow t + \Delta t$
end for

Section 3.5.2 contains the simplest one-dimensional Gaussian case where the optimal Γ is known and it is shown that the algorithm approximates it quickly. A different Gaussian problem extracted from a diffusion bridge context is explored in Section 3.5.3, where the algorithm is shown to approximate a Γ matrix that exhibits an even better empirical asymptotic variance than the one given by Proposition 3.4.5. Finally, the algorithm is applied to finding the optimal Γ in estimating the posterior mean in a Bayesian inference problem in Section 3.5.4, where the situation is shown to be similar to Proposition 3.4.8, in the sense that the optimal Γ is close to 0; after and separately from such a finding, the empircal asymptotic variance for a small Γ is compared that for $\Gamma = I_n$, with dramatic (about tenfold) improvement in both the full gradient and minibatch gradient cases.

3.5.1 Methodology

Here we describe an on-the-fly procedure[1] to repeatedly calculate the change (3.1.8) in Γ by simulating the first variation process parallel to underdamped Langevin processes. The discretisation schemes used to simulate (3.1.1) and (3.3.2) are given in Section 3.5.1. Two gradient procedures, namely gradient descent and the Heavy ball method, for evolving Γ given a gradient are detailed in Section 3.5.1. Then iterates from Section 3.5.1 are used to approximate each change in Γ in Section 3.5.1 (see also Section 3.5.5). The key idea linking the above is that if equation (3.3.4) holds, then

$$\begin{aligned}\Delta\Gamma &= \int \nabla_p\phi \otimes \nabla_p\tilde{\phi}\; d\tilde{\pi}\\ &= -\int \bigg(\int_0^\infty \mathbb{E}[\nabla f(q_s)^\top D_p q_s]^\top ds\bigg)\bigg(\int_0^\infty \mathbb{E}[\nabla f(\tilde{q}_t)^\top D_p\tilde{q}_t]dt\bigg) d\tilde{\pi}, \end{aligned} \tag{3.5.1}$$

where (q_t, p_t) and $(\tilde{q}_t, \tilde{p}_t)$ denote the solutions to (3.1.1) with initial values (q, p), $(q, -p)$ respectively, $(D_p q_t, D_p p_t)$ and $(D_p\tilde{q}_t, D_p\tilde{p}_t)$ denote the solutions to (3.5.3) with $\tilde{q}_t$ replacing q_t for the latter and the integral in (3.5.1) is with respect to (q, p).

Splitting

We split the dynamics (3.1.1) by a so-called *BAOAB* splitting scheme, see [117, 118], in order to integrate the Langevin dynamics (3.1.1). This is given explicitly by

$$\begin{cases} p^{i+\frac{1}{3}} &= p^i - \nabla U(q^i)\frac{\Delta t}{2}\\ q^{i+\frac{1}{2}} &= q^i + p^{i+\frac{1}{3}}\frac{\Delta t}{2}\\ p^{i+\frac{2}{3}} &= \exp(-\Delta t\Gamma^i)p^{i+\frac{1}{3}} + \sqrt{1-\exp(-2\Delta t\Gamma^i)}\xi^i\\ q^{i+1} &= q^{i+\frac{1}{2}} + p^{i+\frac{2}{3}}\frac{\Delta t}{2}\\ p^{i+1} &= p^{i+\frac{2}{3}} - \nabla U(q^{i+1})\frac{\Delta t}{2} \end{cases} \tag{3.5.2}$$

for $i \in \mathbb{N}$, $\Delta t > 0$, where ξ^i are independent n-dimensional standard normal random variables and $\Gamma^i \in \mathbb{S}^n_{++}$ are a sequence of friction matrices to be updated throughout the duration of the algorithm. We mention again recent developments, e.g. [41, 47, 65, 141, 173, 177], on discretisations of the underdamped Langevin dynamics; the majority of the numerical error involved in updating Γ is expected to come from the small number of particles in approximating the integrals in the expression (3.1.8) for $\Delta\Gamma$, so that no further deliberation is made about the choice of discretisation for the purposes here. The

[1] meaning that this procedure occurs simultaneously to the MCMC procedure to approximate $\pi(f)$

first variation process (3.5.3) together with its initial condition is

$$D_p q_t = \int_0^t D_p p_s ds, \tag{3.5.3a}$$

$$D_p p_t = I_n - \int_0^t (D^2 U(q_s) D_p q_s + \Gamma D_p p_s) ds. \tag{3.5.3b}$$

In order to simulate (3.5.3), an analogous splitting scheme is used:

$$\begin{cases} Dp^{i+\frac{1}{3}} & = Dp^i - D^2U(q^i)Dq^i \frac{\Delta t}{2} \\ Dq^{i+\frac{1}{2}} & = Dq^i + Dp^{i+\frac{1}{3}} \frac{\Delta t}{2} \\ Dp^{i+\frac{2}{3}} & = \exp(-\Delta t \Gamma^i) Dp^{i+\frac{1}{3}} \\ Dq^{i+1} & = Dq^{i+\frac{1}{2}} + Dp^{i+\frac{2}{3}} \frac{\Delta t}{2} \\ Dp^{i+1} & = Dp^{i+\frac{2}{3}} - D^2U(q^{i+1})Dq^i \frac{\Delta t}{2}. \end{cases} \tag{3.5.4}$$

In the case where the second derivatives of U are not directly available, the k^{th} column of (for example) $D^2U(q^i)Dq^i \frac{\Delta t}{2}$ can be approximated by

$$-\nabla U\left(q^i + \frac{\Delta t}{2}(Dq^i)_k\right) + \nabla U(q^i) \tag{3.5.5}$$

where $(Dq^i)_k$ denotes the k^{th} column of Dq^i, so that (3.5.3) can still be approximated in the absence of Hessian evaluations. The approximation (3.5.5) will be used only when explicitly stated in the sequel.

Gradient procedure in Γ

Suppose we have available a series of proposal updates $(b_0, \ldots, b_{L-1}) \in \mathbb{R}^{n \times n \times L}$ for Γ, each element of which being noisy estimates of the same gradient direction in Γ. Given stepsizes $\alpha^i = \alpha \in \mathbb{R}$ and an annealing factor $r \in \mathbb{R}$, the following constrained stochastic gradient descent (for i where proposal updates are produced)

$$\Gamma^{i+1} = \Pi^{\mu}_{\text{pd}}\left(\Gamma^i + \frac{\alpha^i}{2L}\sum_{j=0}^{L-1}(b_j + b_j^{\top})\right) \tag{3.5.6}$$

can be considered, where $L \in \mathbb{N}$ and Π^{μ}_{pd} is the projection to a positive definite matrix, for some minimum value $\mu > 0$ that we choose arbitrarily in order to ensure ergodicity, given by

$$\Pi^{\mu}_{\text{pd}}(M) = \sum_{i=1}^{n} \max(\lambda_i, \mu) v_i v_i^{\top} \tag{3.5.7}$$

for symmetric $M \in \mathbb{R}^{n\times n}$ and its the eigenvalue decomposition

$$M = \sum_{i=1}^{n} \lambda_i v_i v_t^\top.$$

Alternatively, a Heavy-ball method [162, 75] (with projection) can be used. The method is considered in the stochastic gradient context in [40], given here as

$$\begin{cases} \Gamma^{i+1} &= \Pi^{\mu}_{\mathrm{pd}}(\Gamma^i + \alpha^i \Theta^{i+1}), \\ \Theta^{i+1} &= (1-\alpha^i r)\Theta^i + \frac{\alpha^i}{2L}\sum_{j=0}^{L-1}(b_j + b_j^\top). \end{cases} \tag{3.5.8}$$

The heavy-ball method offers a smoother trajectory of Γ over the course of the algorithm. Under appropriate assumptions on b_j, in particular if

$$\frac{1}{2L}\sum_{j=0}^{L-1}(b_j + b_j^\top) \sim \mathcal{N}(\nabla\sigma^2(\Gamma^i), \sigma_b^2 I_{n^2}),$$

for some gradient $\nabla\sigma^2(\Gamma^i_k)$ and variance $\sigma_b^2 > 0$, then the system (3.5.8) has the interpretation of an Euler discretisation of a constrained Langevin dynamics, in which case $\frac{r}{\sqrt{\alpha^i}\sigma_b^2}$ is the inverse temperature. By increasing r, the analogous invariant distribution 'sharpens' around the maximum in its density and in this way reduces the effect of noise at equilibrium; on the other hand, decreasing r reduces the decay in the momentum.

A thinning approach for $\Delta\Gamma$

The most straightforward way of approximating the integral in (3.5.1) is to use independent realisations of (3.5.2), as described at the end of Section 3.5.5, but we draw alternatively a thinned sample [153] from a single trajectory here in order to run only a single parallel set of realisations of (3.5.2) and (3.5.4) at a time. More specifically, we consider a single realisation of (3.5.2) and regularly-spaced points from its trajectory (possibly after a burn-in) as sample points from $\tilde{\pi}$. Starting at each of these sample points and ending at each subsequent one, the process is replicated albeit starting with a momentum reversal and simulated in parallel. In addition, for each of the two processes, a corresponding first variation process (3.5.4) is calculated in parallel. A precise description follows.

Let $K = 1$ for simplicity. The Γ direction (3.5.1) is approximated by

$$-\frac{1}{(L+L^*)}\sum_{l=0}^{L+L^*-1}\left(\sum_{i=1}^{T}\frac{\Delta t}{K}\sum_{k=1}^{K}\nabla f(q_{(k)}^{i+Tl+B})^\top Dq_{(k)}^{i+Tl+B}\right)\otimes \left(\sum_{i=1}^{T}\frac{\Delta t}{K}\sum_{k=1}^{K}\nabla f(\tilde{q}_{(k)}^{i+Tl+B})^\top D\tilde{q}_{(k)}^{i+Tl+B}\right), \tag{3.5.9}$$

where $L \in \mathbb{N}$, $((q^i_{(k)}, p^i_{(k)}))_{i\in\mathbb{N}}$, $((\tilde{q}^i_{(k)}, \tilde{p}^i_{(k)}))_{i\in\mathbb{N}}$ denote solutions to (3.5.2)

- for $i \neq B + Tl - 1$, $l \in \mathbb{N}$ if $k \neq 1$ and
- for all i if $k = 1$

with initial condition $(0, 0)$, noise $\xi^i = \xi^i_{(k)}, \tilde{\xi}^i_{(k)}$ for all $i \in \mathbb{N}$ satisfying $\xi^i_{(k)} = \xi^i_{(k')} = \tilde{\xi}^i_{(k)} = \tilde{\xi}^i_{(k')}$ for all $i < B, 1 \leq k \leq K$, $1 \leq k' \leq K$, independent otherwise as i and k vary, along with corresponding $(Dq^i_{(k)}, Dp^i_{(k)})$, $(D\tilde{q}^i_{(k)}, D\tilde{p}^i_{(k)})$ satisfying (3.5.4) for $i \neq B+Tl-1$, $l \in \mathbb{N}_0$ (regardless of k), and where the $k \neq 1$ processes are 'reset' at $i = B+Tl$ corresponding to the values of the $k = 1$ chain if the first variation processes have converged to zero, that is,

$$q_{(k)}^{Tl+B} = q_{(1)}^{Kl+B}, \qquad p_{(k)}^{Tl+B} = p_{(1)}^{Tl+B}, \qquad Dq_{(k)}^{Tl+B} = 0, \qquad Dp_{(k)}^{Tl+B} = I_n \tag{3.5.10a}$$

$$\tilde{q}_{(k)}^{Tl+B} = q_{(1)}^{Tl+B}, \qquad \tilde{p}_{(k)}^{Tl+B} = -p_{(1)}^{Tl+B}, \qquad D\tilde{q}_{(k)}^{Tl+B} = 0, \qquad D\tilde{p}_{(k)}^{Tl+B} = I_n \tag{3.5.10b}$$

for all $1 \leq k \leq K$ if for some $D_{\text{conv}} > 0$,

$$\max_{i,j,k}\left|(Dq_{(k)}^{Tl+B})_{ij}\right| < D_{\text{conv}}, \quad \max_{i,j,k}\left|(D\tilde{q}_{(k)}^{Tl+B})_{ij}\right| < D_{\text{conv}}, \tag{3.5.11a}$$

$$\max_{i,j,k}\left|(Dp_{(k)}^{Tl+B})_{ij}\right| < D_{\text{conv}}, \quad \max_{i,j,k}\left|(D\tilde{p}_{(k)}^{Tl+B})_{ij}\right| < D_{\text{conv}} \tag{3.5.11b}$$

and $L^* \in \mathbb{N}$ is such that the number of elements in $\{l \in \mathbb{N} : 1 \leq l \leq L + L^*\}$ satisfying (3.5.11) is L. The approach is summarised in Algorithm 2. Of course, the above for generic $K \in \mathbb{N}$ constitutes improving approximations to $\Delta\Gamma$. Note that as Γ changes through the prescribed procedure, the asymptotic variance associated to the given observable f is expected to improve, but on the contrary, the estimator (3.5.9) for the continuous-time expression (3.5.1) may well worsen, since the integrand (of the outermost integral) in (3.5.1) is not f. Increasing L is expected to solve any resulting issues; on the other hand extremely small L have been successful in the experiments here.

Remark 3.5.1. If it is of interest to approximate expectations of $P \in \mathbb{N}$ observables with respect to π, the quantity $\sum_i^P \sigma_i^2$ for example can be used as an objective function, where σ_i^2 is the asymptotic variance from the i^{th} observable. In the implementation

Algorithm 2 Gradient procedure in Γ

Result: $\Gamma^i, 1 \le i \le N+1$;
Start from arbitrary $(q^0, p^0) \in \mathbb{R}^{2n}$ and set $Dq^0 = D\tilde{q}^0 = 0$, $Dp^0 = D\tilde{p}^0 = I_n$, $\zeta = \tilde{\zeta} = 0$, $k = 0$, $\Gamma^j = I_n \quad \forall 1 \le j \le B$
for $i = 1 : B-1$ **do**
  compute q^{i+1} according to (3.5.2);
end for
if $i = B$ **then**
  set $(\tilde{q}^i, \tilde{p}^i) \leftarrow (q^i, -p^i)$
end if
for $i = B : N$ **do**
  compute q^{i+1} and $\tilde{q}^{i+1}$ according to (3.5.2);
  compute Dq^{i+1}, $D\tilde{q}^{i+1}$ from (3.5.4) corresponding to q^{i+1}, $\tilde{q}^{i+1}$ respectively;
  compute the row vectors $\quad \zeta \leftarrow \zeta + \nabla f(q^{i+1})^\top Dq^{i+1}\Delta t;$

$$\tilde{\zeta} \leftarrow \tilde{\zeta} + \nabla f(\tilde{q}^{i+1})^\top D\tilde{q}^{i+1}\Delta t$$

  if $l := i - B \in T\mathbb{N}$ and (3.5.11) hold (ignoring appearances of (k)) **then**
    save the matrix $b_{(\frac{k}{G} - \lfloor \frac{k}{G} \rfloor)G} = -\zeta \otimes \tilde{\zeta}$;
    reset as follows: $\quad \zeta, \tilde{\zeta} \leftarrow 0, \qquad (\tilde{q}^{i+1}, \tilde{p}^{i+1}) \leftarrow (q^{i+1}, -p^{i+1});$

$$Dq^{i+1}, D\tilde{q}^{i+1} \leftarrow 0, \qquad Dp^{i+1}, D\tilde{p}^{i+1} \leftarrow I_n$$

    and update the counter $k \leftarrow k+1$;
  end if
  if $k \in G\mathbb{N}$ **then**
    compute Γ^{i+1} according to (3.5.8);
    set $\Gamma^{i+1} = \Gamma^i$.
  end if
end for

in Algorithm 2, instead of only the vectors ζ, $\tilde{\zeta}$, this amounts to calculating at each iteration the vectors $\zeta^{(i)}$, $\tilde{\zeta}^{(i)}$ corresponding to the i^{th} observable and taking the sum of the resulting update matrices in Γ to update Γ. This calls for no extra evaluations of ∇U over the single observable case.

Remark 3.5.2. *(Tangent processes along random directions)* We mention here the situation where simulating the full first variation processes $(D_p q_t, D_p p_t)$ in $\mathbb{R}^{n\times 2n}$ is prohibitively expensive, namely when n^2 is large. In order to calculate changes in Γ, a directional tangent process can be used instead of $(D_p q_t, D_p p_t)$. Consider for a randomly chosen vector $v \in \mathbb{R}^n$ with $|v| = 1$, the pair of vectors $(D_p q_t v, D_p p_t v) \in \mathbb{R}^{n\times 2}$. Multiplying on the right of ((3.5.3) and) (3.5.4) by v, one obtains the system

$$\begin{cases} Dpv^{i+\frac{1}{3}} & = Dpv^i - D^2U(q^i)Dqv^i\frac{\Delta t}{2} \\ Dqv^{i+\frac{1}{2}} & = Dqv^i + Dpv^{i+\frac{1}{3}}\frac{\Delta t}{2} \\ Dpv^{i+\frac{2}{3}} & = \exp(-\Delta t\Gamma^i)Dpv^{i+\frac{1}{3}} \\ Dqv^{i+1} & = Dqv^{i+\frac{1}{2}} + Dpv^{i+\frac{2}{3}}\frac{\Delta t}{2} \\ Dpv^{i+1} & = Dpv^{i+\frac{2}{3}} - D^2U(q^{i+1})Dqv^i\frac{\Delta t}{2}, \end{cases} \tag{3.5.12}$$

where the first term involving the Hessian of U in (3.5.12) can be approximated by

$$-\nabla U(q^i + \frac{\Delta t}{2}Dqv^i) + \nabla U(q^i)$$

and similarly for the last such term. The advantage of this approximation is that it is no longer necessary to work with D^2U, which is an n-by-n matrix, and instead only with ∇U, a length-n vector. In continuous time, the resulting direction in Γ is $\int \nabla\phi^\top v\nabla\tilde{\phi}^\top v d\tilde{\pi} v\otimes v$ and from (3.3.1) the rate of change in asymptotic variance in this direction is $-2(\int \nabla\phi^\top v\nabla\tilde{\phi}^\top v d\tilde{\pi})^2$. However, the trade-off is that the resulting gradient procedure in Γ turns out to be very slow to converge in high dimensions in comparison to simulating a full first variation process; it is illustrative to think of the situation where the randomly chosen vector v is taken from the set of standard Euclidean basis vectors, where only one diagonal value in Γ is changed at a time. See also [187, 82] for such directional derivatives under a different context.

3.5.2 One dimensional quadratic case

Here the algorithm given in Section 3.5.1 is used in the simplest one dimensional

$$U(q) = \frac{V_0}{2}q^2, \qquad f(q) = \frac{1}{2}q^2, \tag{3.5.13}$$

for $V_0 > 0$, case to find the optimal constant friction. Since commutativity issues disappear in the one-dimensional case, the optimal constant friction is known analytically and is given by Proposition 3.4.5 to be $\Gamma = \sqrt{V_0}$, with the asymptotic variance $V_0^{-\frac{5}{2}}$. Moreover, the relationship between the asymptotic variance and Γ is explicitly given by equations (3.4.8) and (3.4.12), which reduces in this case to

$$\sigma^2(\Gamma) = \frac{1}{4V_0^2}\left(\Gamma^{-1} + \frac{1}{V_0}\Gamma\right).$$

The case $V_0 = 5$ is illustrated in Figure 3.5.3. In the middle and right plot of Figure 3.5.3, the procedure in Section 3.5.1 is used for $5 \cdot 10^4$ epochs, with $\Delta t = 0.08$, block-size $T = 125$, $L = 1$ and $D_{\text{conv}} = 2 \cdot 10^{-4}$. Changing the observable to the linear

$$f(q) = q \tag{3.5.14}$$

gives that the 'optimal' (but unreachable in the algorithm due to the constraints) friction is 0 by Corollary 3.4.8. The right plot in Figure 3.5.3 shows that the procedure arrives at a similar conclusion in the sense that the Γ hits and stays at $\mu = 0.2$.

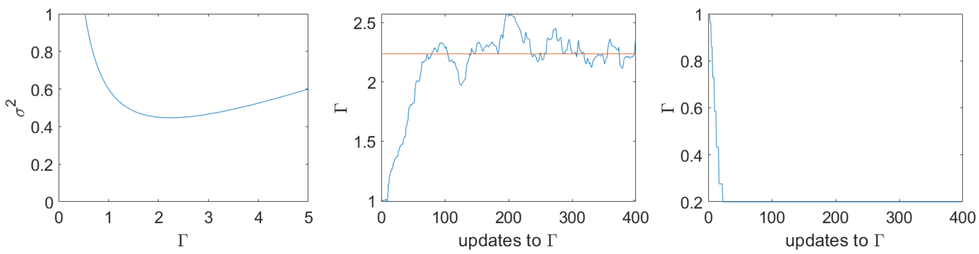

Fig. 3.5.1: Left: Relationship between asymptotic variance and Γ for (3.5.13). Middle and right: Trajectory of Γ for (3.5.13) and (3.5.14) respectively by (3.5.8) with $\alpha^i = 1$, $G = 1$, $r = 0.5$ and $\mu = 0.2$. Middle: the red line is the optimal value $\Gamma = \sqrt{5}$ given by Proposition 3.4.5. All plots are for $V_0 = 5$.

3.5.3 Diffusion bridge sampling

The algorithm in Section 3.5.1 is applied in the context of diffusion bridge sampling [86, 88] (see also for example [13, 48, 87]), where the SDE

$$dx_t = -\nabla V(x_t)dt + \sqrt{2\beta^{-1}}dW_t' \tag{3.5.15}$$

for a suitable $V : \mathbb{R}^d \to \mathbb{R}$, $\beta > 0$ and W_t' standard Wiener process on $\mathbb{R}^d$, is conditioned on the events

$$x_0 = x_- \quad \text{and} \quad x_T = x_+ \tag{3.5.16}$$

for some fixed $T > 0, x_0, x_+ \in \mathbb{R}^d$ and the problem setting is to sample from the path space of solutions to (3.5.15) conditioned on (3.5.16). For the derivation of the following formulations, we refer to Section 5 in [86] and Section 6.1 in [12]; here we extract a simplified potential U to apply our algorithm on after a brief description.

Let

$$V(x) = \frac{1}{2}|x|^2, \qquad x_- = x_+ = 0, \qquad \beta = 1, \qquad d = 1, \qquad T = 1.$$

Using the measure given by Brownian motion conditioned on (3.5.16) as the reference measure μ_0 on the path space of continuous functions $C([0,1],\mathbb{R})$, the measure μ associated to (3.5.15) conditioned on (3.5.16) satisfies $\frac{d\mu}{d\mu_0}(x) \propto \exp\left(-\frac{1}{4}\int_0^T |x|^2 dt\right)$, where the left hand side denotes the Radon-Nikodym derivative, so that discretising μ on a grid in $[0,1]$ with grid-size $\delta > 0$ gives the approximating measure $\pi(q_1,\dots,q_n) \propto e^{-U(q_1,\dots,q_n)}$ where U is given by

$$U(q) = \frac{1}{2}q^\top \Sigma^{-1} q = \frac{1}{2}q^\top \begin{pmatrix} \frac{2}{\delta}+\frac{\delta}{4} & -\frac{1}{\delta} & & & \\ -\frac{1}{\delta} & \frac{2}{\delta}+\frac{\delta}{4} & -\frac{1}{\delta} & & \\ & \dots & & & \\ & & -\frac{1}{\delta} & \frac{2}{\delta}+\frac{\delta}{4} & -\frac{1}{\delta} \\ & & & \frac{2}{\delta}+\frac{\delta}{4} & -\frac{1}{\delta} \end{pmatrix} q.$$

From here the Langevin system (3.1.1) can be used to sample from π and the algorithm given in Section 3.5.1 is applied. For this purpose, the observable $f(q) = \frac{1}{2}|q|^2$ is used together with the parameters $\delta = \frac{1}{21}$, $n = 20$, $K = 1$, $L = 5$, $T = 60$, $B = 100$ and $D_{\text{conv}} = 0.01$. Only the diagonal values of Γ are updated and their trajectories are shown in Figure 3.5.2.

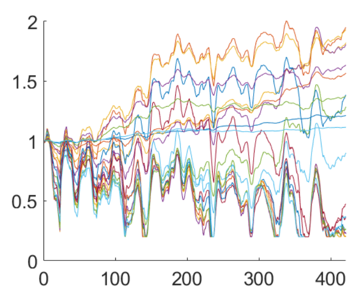

Fig. 3.5.2: Diagonal values of Γ over iterations of (3.5.8) with $\alpha^i = 0.2$, $G = 5$, $r = 1$ and $\mu = 0.2$.

At the end of 300000 epochs, Γ is given by

$$\Gamma_{\text{final}} = \text{diag}(1.2129, 1.5673, 1.8199, 1.8055, 1.2858, 0.9013, 0.3588, 0.2631, \\ 0.2000, 0.2000, 0.2252, 0.2579, 0.3621, 0.4715, 1.3842, 1.9467, \\ 1.9289, 1.6326, 1.3730, 1.1153).$$

This Γ is fixed and used for a standard sampling procedure for the same potential and observable. The asymptotic variance is approximated by grouping the epochs after $B =$

100 burn-in iterations into $N_B = 999$ blocks of $T = 300$ epochs, specifically,

$$\sigma_{\text{approx}} = \frac{1}{N_B} \sum_{l=0}^{N_B-1} \left[\frac{1}{\sqrt{T\Delta t}} \sum_{i=1}^{T} \Delta t \bigg(f(q^{i+Tl+B}) - \frac{1}{N} \sum_{j=1}^{N} \Delta t f(q^{j+B}) \bigg) \right]^2,$$

where q^i are iterates in the numerical approximation of q_t, and this is compared to the estimate from the same procedure using different values of fixed Γ in Table 3.5.1. Note that $\Gamma = \Sigma^{-\frac{1}{2}}$ is the optimal Γ in the restricted class of matrices commuting with Σ given by Proposition 3.4.5, where the asymptotic variance is known to be $\mathrm{Tr}(\Sigma^{\frac{5}{2}}) \approx 6.4785$.

	σ_{approx}
$\Gamma = I_n$	6.9834
$\Gamma = \Sigma^{-\frac{1}{2}}$	6.5096
$\Gamma = \Gamma_{\text{final}}$	6.1667

Tab. 3.5.1: Empirical asymptotic variances with $N_B = 999$, $T = 300$, $B = 100$, $N = 299700$.

3.5.4 Bayesian inference

We adopt the binary regression problem as in [58] on a dataset[2] with datapoints encoding information about images on a webpage and each labelled with 'ad' or 'non-ad'. The labels $\{Y_i\}_{1\leq i\leq p}$, taking values in $\{0,1\}$, of the $p = 2359$ datapoints (counting only those without missing values) given in the dataset are modelled as conditionally independent Bernoulli random variables with probability $\{\rho(\beta^\top X_i)\}_{1\leq i\leq p}$, where ρ is the logistic function given by $\rho(z) = e^{cz}/(1+e^{cz})$ for all $z \in \mathbb{R}$, $c \in \mathbb{R}$ is given by (3.5.18), $\{X_i\}_{1\leq i\leq p}$, β, both taking values in $\mathbb{R}^n$, are respectively vectors of known features from each datapoint and regression parameters to be determined. The parameters β are given the prior distribution $\mathcal{N}(0,\Sigma)$, where $\Sigma^{-1} = \frac{1}{p}\sum_{i=1}^{p} X_i^\top X_i \in \mathbb{R}^{n\times n}$, and the density of the posterior distribution of β is given up to proportionality by

$$\pi_\beta(\beta | \{(X_i,Y_i)\}_{1\leq i\leq p}) \propto \exp\left(\sum_{i=1}^{p} \{cY_i\beta^\top X_i - \log(1+e^{c\beta^\top X_i})\} - \frac{1}{2}\beta^\top \Sigma^{-1}\beta \right),$$

so that the log-density gradient, in our notation $-\nabla U$, is given by

$$-\nabla U(\beta) = \sum_{i=1}^{p} cX_i(Y_i - (1+e^{-c\beta^\top X_i})^{-1}) - \Sigma^{-1}\beta.$$

[2] http://archive.ics.uci.edu/ml/datasets/Internet+Advertisements. Note that besides missing values at some datapoints, the dataset comes with many quantitatively duplicate features and also some linear dependence between the vectors made up of a single feature across all datapoints; here features have been removed so that the said vectors remaining are linearly independently. In particular, $n = 642$.

The observable vector $f_i(q) = q_i$, $1 \le i \le n$, corresponding to the posterior mean is used. The coordinate transform $\hat{\beta} = \Sigma^{-\frac{1}{2}}\beta$ is made before applying the symmetric preconditioner $\Sigma^{\frac{1}{2}}$ on the Hamiltonian part of the dynamics so that the dynamics simulated are as in (3.1.1) with $M = I_n$ and

$$-\nabla U(\hat{\beta}) = \Sigma^{\frac{1}{2}} \sum_{i=1}^{p} cX_i(Y_i - (1 + e^{-c(\Sigma^{\frac{1}{2}}\hat{\beta})^\top X_i})^{-1}) - \hat{\beta}. \tag{3.5.17}$$

We use the observable vector $f_i(\hat{\beta}) = \hat{\beta}_i$, $1 \le i \le n$ and the sum of their corresponding asymptotic variances as the value to optimise with respect to Γ, but show in Figures 3.5.3 and 3.5.4 the estimated asymptotic variances for both sets $f_i(\hat{\beta})$, $f_i(\beta)$ of observables, where the estimation is calculated using the vector on the left of the outer product in (3.5.9) in accordance with $2\int \nabla\phi^\top \Gamma \nabla\phi d\tilde{\pi}$ which follows from the formula (3.2.12) after integrating by parts with truncation. The approximation (3.5.5) for the term(s) including the Hessian in (3.5.4) has been used to test the method despite the explicit availability of the Hessian. During the execution of Algorithm 2, the constant c has been set to

$$c = \bar{c} := \frac{5}{\max_i(\Sigma^{\frac{1}{2}} \sum_j X_j Y_j)_i}. \tag{3.5.18}$$

In detail, 30000 epochs are simulated; after 100 burn-in iterations of the Langevin discretisation (3.5.2), 2 parallel simulations of (3.5.2) and 2 of the first variation discretisation (3.5.4) are run according to Section 3.5.1 with time-step $\Delta t = 0.1$, block-size $T = 100$, $L = 1$ block per update in Γ, $K = 1$ and tolerance $D_{\text{conv}} = 0.01$.

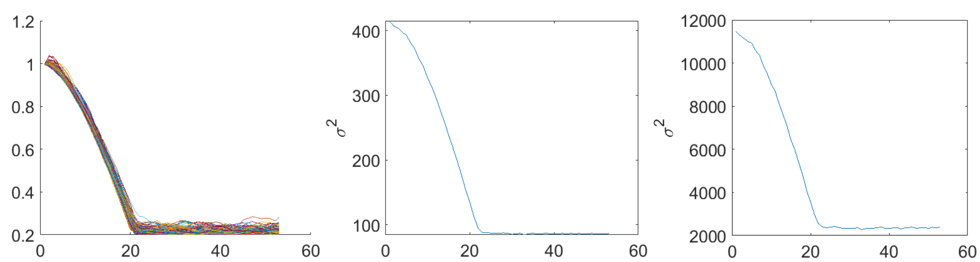

Fig. 3.5.3: Left: Diagonal values of Γ over iterations of (3.5.8) with $\alpha^i = 0.1$, $G = 1$, $r = 1$ and $\mu = 0.2$. Note that the mean of the absolute values of all entries of Γ at the end of the iterations is 0.0039. Middle: Sum over i of estimated asymptotic variances for $f_i(\hat{\beta})$; right: for $f_i(\beta)$.

In Figures 3.5.3 and 3.5.4, Γ starts initially from the identity I_n and descends towards $0.2I_n$ (restricted as in (3.5.7)), as expected for a linear observable and potential close to a quadratic (see Proposition 3.4.9). We note that in the gradient descent procedure for Γ, using the minibatch gradient does not change the behaviour shown in Figures 3.5.3

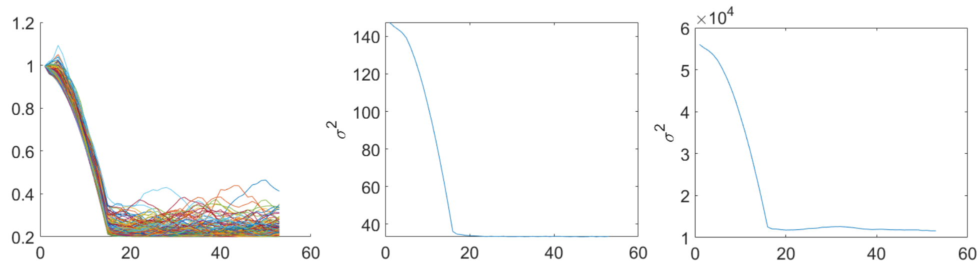

Fig. 3.5.4: The same as in the caption of Figure 3.5.4, except $r = 0.5$ and a different dataset (https://archive.ics.uci.edu/ml/datasets/Musk+(Version+1)) is used where $n = 167$ and $p = 476$. The mean of the absolute values of all entries of Γ at the end of the iterations is 0.0210.

and 3.5.4. In addition, although the trajectory of Γ seems to go directly to zero, we expect the optimal Γ to be close but away from zero since the potential is close but not exactly quadratic.

Next, the value for Γ is fixed at various values and used for hyperparameter training on the same problem for the first dataset, using both the full gradient (3.5.17) and a mini-batch[3] version where the sum in (3.5.17) is replaced by $\frac{p}{10}$ times a sum over a subset S of $\{1,\dots,p\}$ with 10 elements randomly drawn without replacement such that S changes once for each i in (3.5.2). In the minibatch gradient case, c is set to a fraction of (3.5.18), specifically $\bar{c}(\frac{p}{10})^{-1}$. In Tables 3.5.2 and 3.5.3, variances for the posterior mean estimates are shown (similar variance reduction results persist when using the probability of success for features taken from a single datapoint in the dataset).

In detail, for each row of Tables 3.5.2 and 3.5.3, $N = 29700$ epochs of (3.5.2) are simulated with the same parameters as above. The asymptotic variance for each observable entry is approximated using block averaging (Section 2.3.1.3 in [123]) by grouping the epochs after $B = 100$ burn-in iterations into $N_B = 99$ blocks of $T = 300$ epochs, that is,

$$\sigma^2_{k,\text{approx}} = \frac{1}{N_B}\sum_{l=0}^{N_B-1}\left[\frac{1}{\sqrt{T\Delta t}}\sum_{i=1}^{T}\Delta t\Big(f_k(q^{i+Tl+B}) - \frac{1}{N}\sum_{j=1}^{N} f_k(q^{j+B})\Big)\right]^2$$

and $N_B = 3$ blocks of $T = 9900$ epochs (respectively for each column of Tables 3.5.2 and 3.5.3); the values 0.8667 and 0.1571 approach and correspond to values in the middle plot of Figure 3.5.3 after multiplying by $n = 642$. The variances are compared to those using a gradient oracle: unadjusted (overdamped) Langevin dynamics[58] and with an

[3] The control variate stochastic gradient on underdamped dynamics [39, 149] is not directly considered here but the benefits of an improved Γ is expected to carry over to such variations of the stochastic gradient.

irreversible perturbation[55], where the antisymmetric matrix J is given by

$$J_{i,j} = \begin{cases} 1 & \text{if } j-i=1 \text{ or } 1-n, \\ -1 & \text{if } i-j=1 \text{ or } 1-n, \\ 0 & \text{otherwise} \end{cases}$$

for $1 \leq i,j \leq n$ and the stepsizes are the same as for underdamped implementations. In addition, the Euclidean distance from intermediate estimates of the posterior mean to a total, combined estimate is shown for each method. Specifically, $d_k := \left| \frac{1}{300k} \sum_{i=1}^{300k} f(q^{i+B}) - \hat{\pi}(f) \right|$ is plotted against k in Figure 3.5.5, where $\hat{\pi}(f)$ is the mean (over the methods listed in Tables 3.5.2 and 3.5.3) of the final posterior mean estimates. A weighted mean with unit weights except one half on the $\Gamma = 0.2I_n$ and $\Gamma = 0.1I_n$ methods also gave similar results, though this is not shown explicitly.

	block-size $T = 300$	block-size $T = 9900$
$\Gamma = I_n$	(1.2669,0.0320)	(0.8667,0.7190)
$\Gamma = 0.2I_n$	(0.2939,0.0018)	(0.1571,0.0243)
$\Gamma = 0.1I_n$	(0.1739,0.0007)	(0.0890,0.0092)
overdamped	(1.2298,0.0319)	(0.8687,0.8662)
irreversible overdamped	(0.5642,0.0077)	(0.3835,0.1614)

Tab. 3.5.2: $(\frac{1}{n}\sum_{k=1}^{n} \sigma^2_{k,\text{approx}}, \frac{1}{n}\sum_{k=1}^{n}(\sigma^2_{k,\text{approx}} - \frac{1}{n}\sum_{l=1}^{n}\sigma^2_{l,\text{approx}})^2)$ - Empirical asymptotic variances, mean and variance over observable entries, where full gradients have been used.

	block-size $T = 300$	block-size $T = 9900$
$\Gamma = I_n$	(1.9575,0.0744)	(1.3338,1.6650)
$\Gamma = 0.2I_n$	(0.4600,0.0042)	(0.2781,0.0784)
$\Gamma = 0.1I_n$	(0.2646,0.0016)	(0.1335,0.0208)
overdamped	(1.9137,0.0791)	(1.3065,1.9714)
irreversible overdamped	(0.8764,0.0150)	(0.5778,0.3266)

Tab. 3.5.3: The same as in Table 3.5.2, except for minibatch gradients

These figures demonstrate improvement of an order of magnitude in observed variances for Γ close to that resulting from the gradient procedure over $\Gamma = I_n$. The improvement is also seen when compared to overdamped Langevin dynamics with and without irrreversible perturbation.

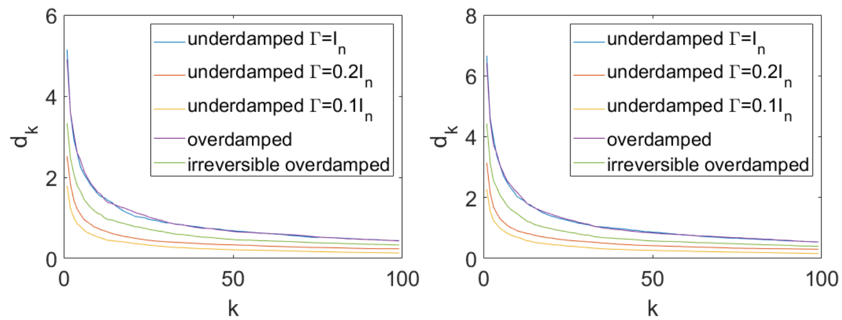

Fig. 3.5.5: Euclidean distances to a combined posterior mean estimate over time. Left: full gradient. Right: minibatch gradient.

3.5.5 Alternative methods

Here we complete the section on numerical methods by detailing alternative ways to approximate the gradient value for our friction optimisation procedure.

Solving the Poisson equation with a Galerkin method

Throughout this section 2.3.1, $M = I_n$ is assumed. In low dimensions, it is feasible to approximate $\nabla_p \phi$ and a change in Γ using Hermite polynomials. This approach gives an approximation in a finite subspace of $L^2(\tilde{\pi})$ at the level of ϕ, as opposed to estimates of $\nabla_p \phi$ at particular points in space as in the Monte Carlo approach in Section 3.5. Specifically, the polynomials given by

$$H_l(z) = \frac{(-1)^l}{\sqrt{l!}} e^{\frac{z^2}{2}} \frac{d^l}{dz^l}\left(e^{-\frac{z^2}{2}}\right)$$

for $l \in \mathbb{N}_0$ and their products in the multidimensional case

$$H_{\underline{l}}(p) = \prod_{k=1}^{n} H_{\underline{l}_k}(p_k), \qquad p = (p_1, \dots, p_n) \in \mathbb{R}^n$$

for multiindices $\underline{l} = (\underline{l}_1, \dots, \underline{l}_n) \in \mathbb{N}_0^n$ are considered in the weighted $L^2(\omega)$ space, where ω is given by $\omega(p) = \frac{e^{-\frac{1}{2}|p|^2}}{(2\pi)^{-\frac{n}{2}}}$. A property of the Hermite polynomials that is repeatedly used here is that

$$\partial_z H_l(z) = \sqrt{l} H_{l-1}(z).$$

For the application of Hermite polynomials in solving the Poisson equation associated to Langevin dynamics (in the case of scalar friction), we refer to [168]. See also Chapter 5 in [84] for Hermite polynomials in the multidimensional setting. In the case of a non-quadratic potential U, the same polynomials are used here after a Gram-Schmidt procedure in $L^2(\pi)$, which are denoted $(\hat{H}_{\underline{l}})_{\underline{l}\in\mathbb{N}_0^n}$, so that

$$\hat{H}_{\underline{l}} = \sum_{|\underline{k}|_\infty \leq K} \alpha_{\underline{l}}^{\underline{k}} H_{\underline{k}},$$

where $|\underline{k}|_\infty = \max(\underline{k}_1, \dots, \underline{k}_n)$, $K \in \mathbb{N}$, for some constants $\alpha_{\underline{k}}^{\underline{l}} \in \mathbb{R}$ calculated numerically. Their products with $H_{\underline{l}}$ are considered on $L^2(\tilde{\pi})$. Similarly, Fourier approximations can be used in the case of an n-torus (in q).
The observable $f \in L_0^2(\pi)$ is approximated by the projection defined by

$$\Pi_K^q f := \sum_{|\underline{l}|_\infty \leq K} \hat{H}_{\underline{l}} \int f \hat{H}_{\underline{l}} d\pi = \sum_{|\underline{k}|_\infty, |\underline{l}|_\infty \leq K} \hat{H}_{\underline{l}} \alpha_{\underline{l}}^{\underline{k}} \int f H_{\underline{k}} d\pi. \tag{3.5.19}$$

Since the generator has the form

$$\mathcal{L} = \nabla_p^* \cdot \nabla_q - \nabla_q^* \cdot \nabla_p - (\nabla_p^*)^\top \Gamma \nabla_p,$$

where

$$\nabla_q^* = -\nabla_q + \nabla U, \qquad \nabla_p^* = -\nabla_p + p$$

are the respective formal $L^2(\tilde{\pi})$-adjoints of ∇_q and ∇_p, the negative of the generator in the Poisson equation applied on functions of the form (3.5.19) is the $(K+1)^{2n}$-by-$(K+1)^{2n}$ matrix given by

$$\begin{aligned} L_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}} &= \langle \hat{H}_{\underline{k}} H_{\underline{l}}, -\mathcal{L}(\hat{H}_{\underline{\hat{k}}} H_{\underline{\hat{l}}}) \rangle_{\tilde{\pi}} \\ &= -\langle \hat{H}_{\underline{k}} \nabla_p H_{\underline{l}}, \nabla_q \hat{H}_{\underline{\hat{k}}} H_{\underline{\hat{l}}} \rangle_{\tilde{\pi}} + \langle \nabla_q \hat{H}_{\underline{k}} H_{\underline{l}}, \hat{H}_{\underline{\hat{k}}} \nabla_p H_{\underline{\hat{l}}} \rangle_{\tilde{\pi}} + \langle \hat{H}_{\underline{k}} \nabla_p H_{\underline{l}}, \Gamma \hat{H}_{\underline{\hat{k}}} \nabla_p H_{\underline{\hat{l}}} \rangle_{\tilde{\pi}} \\ &= -\sum_i \langle \hat{H}_{\underline{k}}, \partial_{q_i} \hat{H}_{\underline{\hat{k}}} \rangle_\pi (\sqrt{\underline{l}_i} \delta_{\underline{\hat{l}}}^{\underline{l}-e_i}) + \sum_i \langle \partial_{q_i} \hat{H}_{\underline{k}}, \hat{H}_{\underline{\hat{k}}} \rangle_\pi (\sqrt{\underline{\hat{l}}_i} \delta_{\underline{\hat{l}}-e_i}^{\underline{l}}) \\ &\quad + \sum_{i,j} \delta_{\underline{\hat{k}}}^{\underline{k}} \delta_{\underline{\hat{l}}-e_i}^{\underline{l}-e_j} \sqrt{\underline{l}_j \underline{\hat{l}}_i} \Gamma_{i,j} \end{aligned} \tag{3.5.20}$$

where δ denotes the Kronecker delta here, the dependences of $\hat{H}_{\underline{k}}$, $\hat{H}_{\underline{\hat{k}}}$ and $H_{\underline{l}}$, $H_{\underline{\hat{l}}}$ on q and p respectively have been suppressed, $\langle v, w \rangle$ denotes $\sum_i \langle v_i, w_i \rangle$ for $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n)$ and $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2(\tilde{\pi})$. Note further that

$$\langle \partial_{q_i} \hat{H}_{\underline{k}}, \hat{H}_{\underline{\hat{k}}} \rangle_\pi = \sum_{|\underline{l}|_\infty, |\underline{\hat{l}}|_\infty \leq K} \alpha_{\underline{k}}^{\underline{l}} \sqrt{\underline{l}_i} \langle H_{\underline{l}-e_i}, H_{\underline{\hat{l}}} \rangle_\pi \alpha_{\underline{\hat{k}}}^{\underline{\hat{l}}},$$

so that since $\alpha_{\underline{k}}^{l}$ are derived from the inner products in $L^2(\pi)$ between the original Hermite polynomials $(H_{\underline{l}})_{\underline{l}}$, these inner products are the only values to be computed numerically other than those for the projection $\Pi_K^q f$ of the observable onto the finite dimensional subspace of $L^2(\tilde{\pi})$ spanned by the first $K+1$ Hermite polynomials given by (3.5.19). Solving the Poisson equation then reduces to finding the coefficients $\phi_{\underline{k},\underline{l}} \in \mathbb{R}$ of

$$\Pi_K^{(q,p)}\phi = \sum_{|\underline{k}|_\infty, |\underline{l}|_\infty \le K} \phi_{\underline{k},\underline{l}} \hat{H}_{\underline{k}} H_{\underline{l}}$$

solving the linear system

$$\sum_{\left|\underline{\hat{k}}\right|_\infty, \left|\underline{\hat{l}}\right|_\infty \le K} L_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}} \phi_{\underline{\hat{k}},\underline{\hat{l}}} = \Pi_K^{(q,p)} f = \begin{cases} \sum_{\left|\underline{\hat{k}}\right|_\infty \le K} \alpha_{\underline{k}}^{\underline{\hat{k}}} \int f H_{\underline{\hat{k}}} d\pi & \text{if } \underline{l} = \underline{0} \\ 0 & \text{otherwise,} \end{cases} \tag{3.5.21}$$

where note $L_{\underline{k},\underline{l},\underline{0},\underline{0}} = L_{\underline{0},\underline{0},\underline{\hat{k}},\underline{\hat{l}}} = 0$ so that only $\phi_{\underline{k},\underline{l}}$ for $(\underline{k},\underline{l}) \neq (\underline{0},\underline{0})$ are determined by (3.5.21) and $\phi_{\underline{0},\underline{0}} = 0$ is enforced independently. Finally, the gradient direction in Γ is given by

$$\begin{aligned}
(\Delta\Gamma)_{i,j} &= \int \sum_{|\underline{k}|_\infty, |\underline{l}|_\infty \le K} \phi_{\underline{k},\underline{l}} \hat{H}_{\underline{k}} \sqrt{\underline{l}_i} H_{\underline{l}-e_i} \sum_{\left|\underline{\hat{k}}\right|_\infty, \left|\underline{\hat{l}}\right|_\infty \le K} \phi_{\underline{\hat{k}},\underline{\hat{l}}} (-1)^{\left|\underline{\hat{l}}\right|} \hat{H}_{\underline{\hat{k}}} \sqrt{\underline{\hat{l}}_j} H_{\underline{\hat{l}}-e_j} d\tilde{\pi} \\
&= \sum_{|\underline{k}|_\infty, |\underline{l}|_\infty, \left|\underline{\hat{k}}\right|_\infty, \left|\underline{\hat{l}}\right|_\infty \le K} \phi_{\underline{k},\underline{l}} \phi_{\underline{\hat{k}},\underline{\hat{l}}} \sqrt{\underline{l}_i \underline{\hat{l}}_j} (-1)^{\left|\underline{\hat{l}}\right|} \delta_{\underline{\hat{k}}}^{\underline{k}} \delta_{\underline{\hat{l}}-e_j}^{\underline{l}-e_i} \\
&= \sum_{|\underline{k}|_\infty, |\underline{l}|_\infty \le K} \phi_{\underline{k},\underline{l}} \phi_{\underline{k},\underline{l}-e_i+e_j} \sqrt{\underline{l}_i (\underline{l}-e_i+e_j)_j} (-1)^{|\underline{l}|}
\end{aligned} \tag{3.5.22}$$

where $\left|\underline{\hat{l}}\right| = \underline{\hat{l}}_1 + \cdots + \underline{\hat{l}}_n$ and $\phi_{\underline{k},\underline{l}} = 0$ if there is some i such that $\underline{k}_i > K$ or $\underline{l}_i > K$. More robustly, the asymptotic variance can be discretised first, followed by taking the gradient direction with respect to the approximate asymptotic variance. Namely, (half of) the asymptotic variance $\int \nabla_p \phi^\top \Gamma \nabla_p \phi d\tilde{\pi}$ can be approximated by

$$\sum_{|\underline{k}|_\infty, |\underline{l}|_\infty, \left|\underline{\hat{k}}\right|_\infty, \left|\underline{\hat{l}}\right|_\infty \le K} \phi_{\underline{k},\underline{l}} L_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}} \phi_{\underline{\hat{k}},\underline{\hat{l}}} \tag{3.5.23}$$

(or simply the last term in (3.5.20) replacing $L_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}}$), so that the derivative with respect to the entries $\Gamma_{i,j}$ can be taken as follows. With abuse of notation, let $L^{-1} \in \mathbb{R}^{(K+1)^{2n}-1 \times (K+1)^{2n}-1}$ be the inverse of the matrix depending on Γ given by (3.5.20) with the $L_{\underline{k},\underline{l},\underline{0},\underline{0}} = L_{\underline{0},\underline{0},\underline{\hat{k}},\underline{\hat{l}}} = 0$ entries removed. Let also $\underline{\phi} \in \mathbb{R}^{(K+1)^{2n}}$ be the vector made up of the coefficients $\phi_{\underline{k},\underline{l}}$ for $\underline{k}+\underline{l} \neq 0$ so that equation (3.5.21) can be rewritten as

$$\underline{\phi} = L^{-1}((\Pi_K^{(q,p)} f)_2, \dots, (\Pi_K^{(q,p)} f)_{(K+1)^{2n}})^\top.$$

By (3.5.20), the derivative of $L_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}}$ with respect to the entry $\Gamma_{i,j}$ is

$$\partial L^{i,j}_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}} := \delta^{\underline{k}}_{\underline{\hat{k}}} \delta^{\underline{l}-e_j}_{\underline{\hat{l}}-e_i} \sqrt{\underline{l}_j \underline{\hat{l}}_i}.$$

Let $\partial L^{i,j} \in \mathbb{R}^{(K+1)^{2n}-1 \times (K+1)^{2n}-1}$ denote the matrix with entries $\partial L^{i,j}_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}}$ except that the $\partial L^{i,j}_{\underline{k},\underline{l},\underline{0},\underline{0}}$ and $\partial L^{i,j}_{\underline{0},\underline{0},\underline{\hat{k}},\underline{\hat{l}}}$ entries are deleted. The derivative of (3.5.23) with respect to the entry $\Gamma_{i,j}$ is then

$$\underline{\phi}^\top \partial L^{i,j} \underline{\phi} + \sum_{|\underline{k}|_\infty, |\underline{l}|_\infty, |\underline{\hat{k}}|_\infty, |\underline{\hat{l}}|_\infty \leq K} 2(\partial \phi^{i,j})_{\underline{k},\underline{l}} L_{\underline{k},\underline{l},\underline{\hat{k}},\underline{\hat{l}}} \phi_{\underline{\hat{k}},\underline{\hat{l}}}, \tag{3.5.24}$$

where $\partial \phi^{i,j} \in \mathbb{R}^{(K+1)^{2n}}$ is the vector given by

$$(\partial \phi^{i,j})_k := \begin{cases} 0 & \text{if } k = 1 \\ -(L^{-1} \partial L^{i,j} \underline{\phi})_{k-1} & \text{otherwise,} \end{cases}$$

so that the gradient direction in Γ is given by the negative of (3.5.24).
It's also possible to approximate ϕ using a finite difference in q, Hermite projection in p approach in the case when the state space in q is the n-torus; we omit further descriptions but refer to [64] for this direction.

Approximation of $\Delta\Gamma$ using independent realisations

One can use the ending values of a number of independent realisations of (3.5.2) to approximate the integral with respect to π in (3.5.1) and, for each of those realisations, to use two additional sets of realisations of (3.5.2) and (3.5.4) to approximate each of the expectations under the integral in (3.5.1). This is alternative to the thinning approach described earlier.
Fix a starting point (q,p); the first of the expectations in (3.5.1) (and similarly for the second) can be approximated at time $s = i\Delta t$ with

$$\frac{1}{K} \sum_{k=1}^{K} \nabla f(q^i_{(k)})^\top D q^i_{(k)}$$

where $K \in \mathbb{N}$, $(q^i_{(k)}, p^i_{(k)})_{i\in\mathbb{N}}$ denotes the solution to (3.5.2) with initial condition (q, p), noise $\xi^i = \xi^i_{(k)}$ for all $i \in \mathbb{N}$ and where $\xi^i_{(k)}$ are independent as $k = 1, \ldots, K$, i varies and $(Dq^i_{(k)}, Dp^i_{(k)})$ is the corresponding solution to (3.5.4). Subsequently, introducing an additional population of independent realisations of (3.5.2) to draw from π after some burn-in period, the change (3.5.1) in Γ can be approximated by

$$-\frac{1}{L}\sum_{l=1}^{L}\bigg(\sum_{i=B+1}^{B+T}\frac{\Delta t}{K}\sum_{k=1}^{K}\nabla f(q^i_{(l,k)})^\top Dq^i_{(l,k)}\bigg)^{\top}\bigg(\sum_{i=B+1}^{B+T}\frac{\Delta t}{K}\sum_{k=1}^{K}\nabla f(\tilde{q}^i_{(l,k)})^\top D\tilde{q}^i_{(l,k)}\bigg)$$

where $B \in \mathbb{N}_0$ is some burn-in number of iterations, $T \in \mathbb{N}$ is some a posteriori number of iterations depending on whether the magnitude of the entries of $Dq^{B+T}_{(l,k)}$ are smaller than some fixed value for all k, l; furthermore $L \in \mathbb{N}$, $((q^i_{(l,k)}, p^i_{(l,k)}))_{i\in\mathbb{N}}$ denotes the solution to (3.5.2) with initial condition say $(0, 0)$, noise $\xi^i = \xi^i_{(l,k)}$ for all $i \in \mathbb{N}$ satisfying

$$\xi^i_{(l,k)} = \xi^i_{(l,k')} \qquad \forall i < B, 1 \le k, k' \le K$$

and are independent otherwise, $((\tilde{q}^i_{(l,k)}, \tilde{p}^i_{(l,k)}))_{i\ge B}$ denotes the solution to (3.5.2) with 'initial' condition

$$(\tilde{q}^B_{(l,k)}, \tilde{p}^B_{(l,k)}) = (q^B_{(l,k)}, -p^B_{(l,k)})$$

for all $1 \le k \le K$, $1 \le l \le L$, independent noise $\xi^i = \tilde{\xi}^i_{(l,k)}$ for $i \ge B$ independent also to $(\xi^i_{(l,k)})_{i\in\mathbb{N}}$. The notation $(Dq^i_{(l,k)}, Dp^i_{(l,k)})$, $(D\tilde{q}^i_{(l,k)}, D\tilde{p}^i_{(l,k)})$ represent the corresponding solutions to (3.5.4).

3.6 Proofs

Theorem 3.6.1. *Let Assumption 5 hold. For any $\mathcal{F}_0$-measurable $z_0 : \Omega \to \mathbb{R}^{2n}$, there exists an almost surely continuous in t solution $(q_t, p_t) = z_t : \Omega \to \mathbb{R}^{2n}$ to* (3.1.1) *that is $\mathcal{F}_t$-adapted and unique up to equivalence. Furthermore, for any $z \in \mathbb{R}^{2n}$, $t \ge 0$, let ρ^z_t be the probability measure given by*

$$\rho^z_t(A) = \mathbb{P}(z^z_t \in A) \tag{3.6.1}$$

for any Borel measurable A, where z^z_t denotes the solution to (3.1.1) *starting at $z_0 = z$, then ρ^z_t*

1. *is a transition probability in the sense that*

   (a) *$(t, z) \mapsto \rho^z_t(A)$ is Borel measurable on $(0, \infty) \times \mathbb{R}^{2n}$,*

   (b) *the Chapman-Kolmogorov relation [67] holds and*

2. *admits a density denoted $\rho(z,\cdot,t):\mathbb{R}^{2n}\to\mathbb{R}$ with respect to the Lebesgue measure on $\mathbb{R}^{2n}$ at every $(t,z)\in(0,\infty)\times\mathbb{R}^{2n}$ such that ρ is a measurable function satisfying for every $z\in\mathbb{R}^{2n}$,*
$$\rho(z,\cdot,\cdot)\in C^\infty(\mathbb{R}^{2n}\times(0,\infty)). \tag{3.6.2}$$

Proof. Theorem 3.5 in [111] together with (3.2.8) yields existence and uniqueness of the solution to (3.1.1). Theorem 3.1 and 3.6 in Section 5 of [67] give that $\rho_t^z(A)$ given by (3.6.1) is a probability kernel, that is, $\rho_t^z(A)$ is Borel measurable in z for fixed A,t, is a probability measure in A for fixed z,t and satisfies the Chapman-Kolmogorov relation. For Borel measurability of $(t,z)\mapsto\rho_t^z(A)$ for fixed A, consider $\hat{z}_t^z$ given by

$$\hat{z}_t^z(\omega)=\begin{cases} z_t^z(\omega) & \text{if } \omega: z_\bullet^z(\omega)\in C([0,\infty)),\\ 0 & \text{otherwise.}\end{cases} \tag{3.6.3}$$

The process $\hat{z}_t^z$ is continuous in t and $\mathcal{F}$-measurable in ω, therefore $\mathbb{P}(\hat{z}_t^z\in A)=\mathbb{P}(z_t^z\in A)$ is continuous in t hence Borel measurable in (t,z). Finally, ρ_t^z admits a density at every $(t,z)\in(0,\infty)\times\mathbb{R}^{2n}$ satisfying (3.6.2) due to Itô's rule and Hörmander's theorem [95]; measurability with respect to the starting point z and therefore jointly in all of the arguments [2, Lemma 4.51] follows by the strong Feller property given by Theorem 4.2 in [51], because $\rho(\cdot,\zeta,t)$ is the pointwise limit of the continuous functions $(\int\eta_k(\zeta-\zeta')\rho(\cdot,\zeta',t)d\zeta')_{k>0}$, where η_k denotes the standard scaled mollifiers. ☐

For all $t\geq 0$, all $z\in\mathbb{R}^{2n}$ and all $f:\mathbb{R}^{2n}\to\mathbb{R}$ integrable under the law $\mathcal{L}((z_t)_{t\geq 0}|z_0=z)$ of z_t starting at z, let

$$P_t(f): z\mapsto\mathbb{E}(f(z_t^z))=\mathbb{E}(f(z_t)|z_0=z). \tag{3.6.4}$$

The family $(P_t)_{t\geq 0}$ forms a strongly continuous (by Proposition 3.6.2) semigroup (by the Markov property; Theorem 3.5 in [111]) on $L^2(\tilde{\pi})$ with unit operator norm. Denote by L the infinitesimal generator associated to this semigroup, given by

$$Lu=\lim_{t\to 0}\frac{P_t(u)-u}{t} \tag{3.6.5}$$

for all functions $u\in\mathcal{D}(L)\subset L^2(\tilde{\pi})$, where the domain $\mathcal{D}(L)$ consists of the functions for which the above limit in $L^2(\tilde{\pi})$ exists.

Proposition 3.6.2. *The family $(P_t)_{t\geq 0}$ is strongly continuous in $L^2(\tilde{\pi})$.*

Proof. Fix $\epsilon>0$. For any $f\in L^2(\tilde{\pi})$, there exists $g\in C_c^\infty$ such that $\|f-g\|_{L^2(\tilde{\pi})}\leq\frac{\epsilon}{3}$.

By triangle inequality, it holds that

$$\|P_t f - f\|_{L^2(\tilde{\pi})} \leq \|P_t f - P_t g\|_{L^2(\tilde{\pi})} + \|f - g\|_{L^2(\tilde{\pi})} + \|P_t g - g\|_{L^2(\tilde{\pi})}. \tag{3.6.6}$$

The last term on the right hand side converges to 0 as $t \to 0$ by Itô's rule. Since the measures $\int \mathbb{E}[\mathbb{1}_{\cdot}(z_t^z)]\tilde{\pi}(dz)$ solve the associated Fokker-Planck equation in the distributional sense, it is equal to the unique solution $\tilde{\pi}$, therefore the first term on the right-hand side of (3.6.6) can be bounded by $\frac{\epsilon}{3}$ after Jensen's inequality. $\square$

Proposition 3.6.3. *The differential operator $-\mathcal{L}$ defined on C_c^∞ has a maximally accretive closure in $L^2(\tilde{\pi})$.*

Proof. Let K denote the differential operator

$$\begin{aligned} K &= e^{-\frac{1}{2}(U(q)+\frac{p^2}{2})}\mathcal{L}(e^{\frac{1}{2}(U(q)+\frac{p^2}{2})}\cdot) \\ &= p^\top M^{-1}\nabla_q - \nabla U(q)^\top \nabla_p + \frac{1}{2}\mathrm{Tr}\Gamma - \frac{1}{4}p^\top \Gamma p + \nabla_p^\top \Gamma \nabla_p \end{aligned}$$

acting on C_c^∞. By a straightforward adaptation of the proof of Proposition 5.5 in [91], the closure of $-K$ in $L^2(\mathbb{R}^{2n})$ and therefore the closure of $-\mathcal{L}$ in $L^2(\tilde{\pi})$ are maximally accretive. $\square$

Proof of Lemma 3.3.1. By Proposition 3.6.3, there are $\phi_k \in C_c^\infty$ such that $(\phi_k, -\mathcal{L}\phi_k)_{k\in\mathbb{N}}$ is an approximating sequence to $(\phi, -L\phi)$ in $L^2(\tilde{\pi})^2$. We have

$$\begin{aligned} \lambda_m \int |\nabla_p \phi_k - \nabla_p \phi_{k'}|^2 d\tilde{\pi} &\leq \int \nabla_p(\phi_k - \phi_{k'})^\top \Gamma \nabla_p(\phi_k - \phi_{k'}) d\tilde{\pi} \\ &= -\int (\phi_k - \phi_{k'})(\mathcal{L}\phi_k - \mathcal{L}\phi_{k'}) d\tilde{\pi}, \end{aligned} \tag{3.6.7}$$

so that $\nabla_p \phi_k$ is Cauchy, with limit denoted as $g \in L^2(\tilde{\pi})$. For any $h \in C_c^\infty$,

$$\left| \int gh + \int \phi \nabla_p h \right| \leq \left| \int gh - \int \nabla_p \phi_k h \right| + \left| \int \phi \nabla_p h - \int \phi_k \nabla_p h \right|,$$

hence

$$\nabla_p \phi_k \to g = \nabla_p \phi \in L^2(\tilde{\pi}). \tag{3.6.8}$$

$\square$

Some additional preliminaries are presented here for the proof of Theorem 3.3.2. For small $\epsilon \in \mathbb{R}$ and some direction $\delta\Gamma \in \mathbb{R}^{n\times n}$ such that $\Gamma + \epsilon\delta\Gamma \in \mathbb{S}_{++}^n$, let L_ϵ be the infinitesimal generator of (3.1.1) with the perturbed friction matrix $\Gamma + \epsilon\delta\Gamma$ in place of Γ, given formally by the differential operator

$$-\mathcal{L}_\epsilon = -p^\top M^{-1}\nabla_q + \nabla U(q)^\top \nabla_p + p^\top M^{-1}(\Gamma + \epsilon\delta\Gamma)\nabla_p - \nabla_p^\top(\Gamma + \epsilon\delta\Gamma)\nabla_p.$$

The formal $L^2(\tilde{\pi})$-adjoint of $\mathcal{L}_\epsilon$ is denoted

$$-\mathcal{L}_\epsilon^* = p^\top M^{-1}\nabla_q - \nabla U(q)^\top \nabla_p + p^\top M^{-1}(\Gamma + \epsilon\delta\Gamma)\nabla_p - \nabla_p^\top(\Gamma + \epsilon\delta\Gamma)\nabla_p$$

just as for $\mathcal{L}^*$.

Proof of Theorem 3.3.2. For $\epsilon \leq \epsilon'$, by Theorem 3.2.3 there exists a solution $\phi + \delta\phi_\epsilon \in L_0^2(\tilde{\pi})$ to the Poisson equation with the perturbed generator $-L_\epsilon(\phi + \delta\phi_\epsilon) = f - \pi(f)$. By Theorem 3.2.4, the directional derivative of $\sigma^2(\Gamma)$ in the direction $\delta\Gamma : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is

$$\frac{1}{2}d\sigma^2.\delta\Gamma = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int \delta\phi_\epsilon f d\tilde{\pi}. \tag{3.6.9}$$

By Proposition 3.6.3, there are $\phi_k, \phi_{k,\epsilon} \in C_c^\infty$ such that $(\phi_k, -\mathcal{L}\phi_k)_{k\in\mathbb{N}}$, $(\phi_{k,\epsilon}, -\mathcal{L}_\epsilon\phi_{k,\epsilon})$ are approximating sequences to $(\phi, f - \pi(f))$, $(\phi + \delta\phi_\epsilon, f - \pi(f))$ respectively in $L^2(\tilde{\pi})^2$. Furthermore, in the same way as in the proof of Lemma 3.3.1 to obtain (3.6.8), it holds that

$$\|\nabla_p\phi_k - \nabla_p\phi\|_{L^2(\tilde{\pi})} + \|\nabla_p\phi_{k,\epsilon} - \nabla_p(\phi + \delta\phi_\epsilon)\|_{L^2(\tilde{\pi})} \to 0 \quad \text{as } k \to \infty. \tag{3.6.10}$$

Using the obvious extension on the notation from (3.1.7),

$$\int(\phi_{k,\epsilon} - \phi_k)(f - \pi(f))d\tilde{\pi} = \int(\phi_{k,\epsilon} - \phi_k)(f - \pi(f) + \mathcal{L}^*\tilde{\phi}_k)d\tilde{\pi} - \int(\phi_{k,\epsilon} - \phi_k)\mathcal{L}^*\tilde{\phi}_k d\tilde{\pi}, \tag{3.6.11}$$

where the first term on the right hand side is negligible as $k \to \infty$ for any fixed ϵ due to $\mathcal{L}^*\tilde{\phi}_k = \widetilde{\mathcal{L}\phi_k}$ and the second term is

$$-\int(\phi_{k,\epsilon} - \phi_k)\mathcal{L}^*\tilde{\phi}_k d\tilde{\pi} = \int(-\mathcal{L}_\epsilon\phi_{k,\epsilon} + \mathcal{L}\phi_k)\tilde{\phi}_k d\tilde{\pi} - \int \epsilon(M^{-1}p - \nabla_p)^\top\delta\Gamma\nabla_p\phi_{k,\epsilon}\tilde{\phi}_k d\tilde{\pi}.$$

Again, the first term on the right hand side is negligible for any fixed ϵ as $k \to \infty$ since both terms in the bracket converge to $\pm(f - \pi(f))$. Integration by parts on the last term gives

$$-\int \epsilon(M^{-1}p - \nabla_p)^\top\delta\Gamma\nabla_p\phi_{k,\epsilon}\tilde{\phi}_k d\tilde{\pi} = -\int \epsilon\nabla_p\phi_{k,\epsilon}^\top\delta\Gamma\nabla_p\tilde{\phi}_k d\tilde{\pi}$$

Collecting the above, for any fixed ϵ, taking $k \to \infty$ and using (3.6.10),

$$\int \delta\phi_\epsilon f d\tilde{\pi} = -\int \epsilon \nabla_p \phi^\top \delta\Gamma \nabla_p(\tilde{\phi} + \delta\tilde{\phi}_\epsilon) d\tilde{\pi}$$

holds. Plugging into (3.6.9), the directional derivative becomes

$$\frac{1}{2} d\sigma^2 . \delta\Gamma = -\lim_{\epsilon \to 0} \int \nabla_p \phi^\top \delta\Gamma \nabla_p(\tilde{\phi} + \delta\tilde{\phi}_\epsilon) d\tilde{\pi}. \tag{3.6.12}$$

From here, for any ϵ, the unwanted term under the limit can be controlled by approximating again with $\tilde{\phi}_{k,\epsilon}$,

$$\begin{aligned}
\lambda_m \int & \left|\nabla_p(\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k)\right|^2 d\tilde{\pi} \\
&\leq \int \nabla_p(\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k)^\top (\Gamma + \epsilon\delta\Gamma) \nabla_p(\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k) d\tilde{\pi} \\
&= \int (\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k)(M^{-1}p - \nabla_p)^\top (\Gamma + \epsilon\delta\Gamma) \nabla_p(\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k) d\tilde{\pi}, \\
&= -\int (\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k) \mathcal{L}_\epsilon^* (\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k) d\tilde{\pi} \\
&= -\epsilon \int (\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k)(M^{-1}p - \nabla_p)^\top \delta\Gamma \nabla_p \tilde{\phi}_k d\tilde{\pi} \\
&= -\epsilon \int \nabla_p(\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k)^\top \delta\Gamma \nabla_p \tilde{\phi}_k d\tilde{\pi} \\
&\leq \epsilon C \int \left(\left|\nabla_p(\tilde{\phi}_{k,\epsilon} - \tilde{\phi}_k)\right|^2 + \left|\nabla_p \tilde{\phi}_k\right|^2 \right) d\tilde{\pi},
\end{aligned}$$

where $\lambda_m = \inf_{0<\epsilon\leq\epsilon'} \lambda_m^\epsilon$, λ_m^ϵ is the smallest eigenvalue of $\Gamma + \epsilon\delta\Gamma$ and $C > 0$ is a constant depending on $\delta\Gamma$ and independent of k. Therefore taking $k \to \infty$ and using (3.6.10) gives

$$\int \left|\nabla_p \delta\tilde{\phi}_\epsilon\right|^2 d\tilde{\pi} \leq \frac{\epsilon C}{\lambda_m - \epsilon C} \int \left|\nabla_p \tilde{\phi}\right|^2 d\tilde{\pi}$$

holds for small enough ϵ and putting into (3.6.12) concludes the proof.

□

Proof of Lemma 3.4.1. Substituting (3.4.3), (3.4.4) and (3.4.1) into the Poisson equation (3.1.5), one obtains

$$\begin{aligned}
-& \begin{pmatrix} 0 & M^{-1} \\ -\Sigma^{-1} & -\Gamma M^{-1} \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \cdot \begin{pmatrix} G_S q + Ep + g \\ E^\top q + H_S p + h \end{pmatrix} - \Gamma : H_S \\
&= \frac{1}{2} q \cdot U_0 q + l \cdot q - \frac{1}{2} U_0 : \Sigma.
\end{aligned}$$

Comparing the constant, first order and second order coefficients in p give respectively the sufficient conditions (3.4.5), (3.4.6) and (3.4.7) as stated. $\square$

Proof of Lemma 3.4.2. Comparing coefficients in q in equation (3.4.5) gives

$$2\Gamma : H_S = U_0 : \Sigma \tag{3.6.13}$$

$$h^\top \Sigma^{-1} = l^\top \tag{3.6.14}$$

$$2E\Sigma^{-1} = U_0 + A_2 \tag{3.6.15}$$

and the same for condition (3.4.6) gives

$$M^{-1}G_S = H_S\Sigma^{-1} + M^{-1}\Gamma E^\top, \tag{3.6.16}$$

$$M^{-1}g = M^{-1}\Gamma h. \tag{3.6.17}$$

Condition (3.6.15) yields (3.4.9). Together with (3.4.7), this gives (3.4.10). From the expression (3.4.10) and by symmetry of U_0 , condition (3.6.13) is in turn satisfied:

$$\begin{aligned}
2\Gamma : H_S &= \Gamma : ((\Sigma U_0 - \Sigma A_2 - 2A_1 M)\Gamma^{-1}) \\
&= \sum_{i,j,k,l} \Gamma_{ji}(\Sigma_{ik}(U_0)_{kl} - \Sigma_{ik}(A_2)_{kl} - (A_1)_{ik}M_{kl})(\Gamma^{-1})_{lj} \\
&= \sum_{i,k} (U_0)_{ki}\Sigma_{ki} = U_0 : \Sigma,
\end{aligned}$$

where symmetry of Σ and M have been used. Substituting (3.4.9) and (3.4.10) into equation (3.6.16) then gives (3.4.8). Equations (3.6.14) and (3.6.17) give the equations (3.4.11) for g and h. $\square$

Proof of Lemma 3.4.3. Denote

$$\bar{G} = \begin{pmatrix} G_S & E \\ E^\top & H_S \end{pmatrix}, \quad \bar{U}_0 = \begin{pmatrix} U_0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \bar{\Sigma} = \begin{pmatrix} \Sigma & 0 \\ 0 & M \end{pmatrix}, \quad \bar{g} = \begin{pmatrix} g \\ h \end{pmatrix}, \quad \bar{l} = \begin{pmatrix} l \\ 0 \end{pmatrix}.$$

Each of ϕ and $f - \pi(f)$ are given by

$$\begin{aligned}
\phi(z) &= \frac{1}{2} z \cdot \bar{G} z - \bar{g} \cdot z - \frac{1}{2}\bar{G} : \bar{\Sigma} \\
f(z) - \pi(f) &= \frac{1}{2} z \cdot \bar{U}_0 z - \bar{l} \cdot z - \frac{1}{2}\bar{U}_0 : \bar{\Sigma}
\end{aligned}$$

for $z = (q,p) \in \mathbb{R}^{2n}$. Substituting into $\sigma^2 = 2\langle \phi, f - \pi(f)\rangle_{\tilde{\pi}}$ gives

$$
\begin{aligned}
&2\int \phi(f - \pi(f))d\tilde{\pi} \\
&\quad = \frac{1}{2}\int (z\cdot \bar{G}z)(z\cdot \bar{U}_0 z)d\tilde{\pi} - \frac{1}{2}\int (z\cdot \bar{G}z)\bar{U}_0 : \bar{\Sigma}d\tilde{\pi} + 2\int (\bar{g}\cdot z)(\bar{l}\cdot z)d\tilde{\pi} \\
&\quad\quad - \frac{1}{2}\int \bar{G} : \bar{\Sigma}(z\cdot \bar{U}_0 z)d\tilde{\pi} + \frac{1}{2}(\bar{G} : \bar{\Sigma})(\bar{U}_0 : \bar{\Sigma}),
\end{aligned}
$$

where

$$
\begin{aligned}
\int (z\cdot \bar{G}z)(z\cdot \bar{U}_0 z)d\tilde{\pi} &= \sum_{i,j,u,v} \bar{G}_{ij}(\bar{U}_0)_{uv}\int z_i z_j z_u z_v d\tilde{\pi} \\
&= \sum_{i,j,u,v} \bar{G}_{ij}(\bar{U}_0)_{uv}\Big(\bar{\Sigma}_{ij}\bar{\Sigma}_{uv} + \bar{\Sigma}_{iu}\bar{\Sigma}_{jv} + \bar{\Sigma}_{iv}\bar{\Sigma}_{ju}\Big) \\
&= (\bar{G} : \bar{\Sigma})(\bar{U}_0 : \bar{\Sigma}) + 2\mathrm{Tr}(\bar{G}\bar{\Sigma}\bar{U}_0\bar{\Sigma}).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
2\int \phi(f - \pi(f))d\tilde{\pi} &= \frac{1}{2}(\bar{G} : \bar{\Sigma})(\bar{U}_0 : \bar{\Sigma}) + \mathrm{Tr}(\bar{G}\bar{\Sigma}\bar{U}_0\bar{\Sigma}) - \frac{1}{2}(\bar{G} : \bar{\Sigma})(\bar{U}_0 : \bar{\Sigma}) \\
&\quad + 2\int (\bar{g}\cdot z)(\bar{l}\cdot z)d\tilde{\pi} \\
&= \mathrm{Tr}(\bar{G}\bar{\Sigma}\bar{U}_0\bar{\Sigma}) + 2\bar{g}\cdot \bar{\Sigma}\bar{l}.
\end{aligned}
$$

□

Proof of Proposition 3.4.4. Let

$$
G = \frac{1}{2}M\Sigma U_0\Gamma^{-1}\Sigma^{-1} + \frac{1}{2}\Gamma U_0\Sigma, \quad E = \frac{1}{2}U_0\Sigma, \quad H = \frac{1}{2}\Sigma U_0\Gamma^{-1}, \tag{3.6.18a}
$$

$$
g = \Gamma\Sigma l, \quad h = \Sigma l \tag{3.6.18b}
$$

so that by Lemma 3.4.2, ϕ given by (3.4.4) is the solution to the Poisson equation (3.1.5) and inserting G, g into (3.4.12) gives

$$
2\langle \phi, f - \pi(f)\rangle_{\tilde{\pi}} = \frac{1}{2}\mathrm{Tr}(M\Sigma U_0\Gamma^{-1}U_0\Sigma + \Gamma U_0\Sigma^2 U_0\Sigma) + 2l^\top \Sigma\Gamma\Sigma l. \tag{3.6.19}
$$

The result follows since $A : B > 0$ for $A, B \in \mathbb{S}^n_{++}$. □

Proof of Proposition 3.4.5. Let $\Sigma = P^\top \Sigma_d P$ be the eigendecomposition of Σ for orthogonal P. Since all symmetric matrices in the set commuting with Σ share eigenvectors with Σ, it suffices to find a unique extremal point of the asymptotic variance with respect to the eigenvalues of Γ, call them $(\lambda_i)_{1\leq i\leq n}$, $\lambda_i \geq 0$. Setting again (3.6.18), ϕ given by (3.4.4) is the solution to the Poisson equation (3.1.5) and the asymptotic variance σ^2 given by (3.2.12) becomes

$$2\langle\phi, f-\pi(f)\rangle_{\tilde{\pi}} = \frac{1}{2}\mathrm{Tr}(\Sigma U_0\Gamma^{-1}U_0\Sigma + \Gamma U_0\Sigma^2 U_0\Sigma), \tag{3.6.20}$$

which reduces to a sum of functions of the form $a_i\lambda_i^{-1}+b_i\lambda_i$, $a_i, b_i > 0$ after diagonalising with P and the result follows. ☐

Proof of Proposition 3.4.6. By Lemma 3.4.2, the solution (3.4.4) to the Poisson equation (3.1.5) is

$$\phi = \left(\frac{U_0\Gamma\Sigma}{4} + \frac{MU_0}{4\Gamma}\right)q^2 + \frac{U_0\Sigma}{2}qp + \frac{U_0\Sigma}{4\Gamma}p^2 + \Sigma\Gamma lq + \Sigma lp - \frac{U_0\Gamma\Sigma^2}{4} - \frac{MU_0\Sigma}{2\Gamma}.$$

By Lemma 3.4.3, the asymptotic variance is given by

$$2\int \phi(f-\pi(f))d\tilde{\pi} = 2\Sigma^2\left(\frac{U_0^2\Sigma}{4} + l^2\right)\Gamma + \frac{U_0^2\Sigma^2}{2\Gamma},$$

which attains the stated minimum at the stated Γ. ☐

Proof of Proposition 3.4.7. Let $\Gamma = \gamma I_n$, $\gamma \in \mathbb{R}$. Note there is a unique solution $\phi \in L_0^2(\tilde{\pi})$ to (3.1.5) by Theorem 3.2.3. The solution ϕ to (3.1.5) has the expression $\phi = \sum_i \alpha_i(\gamma q_i + p_i)$. The asymptotic variance is equal to

$$\begin{aligned}
2\langle\phi, f-\pi(f)\rangle_{\tilde{\pi}} &= 2\gamma\sum_{i,j}\alpha_i\alpha_j\int_{\mathbb{R}^n} q_i\partial_{q_j}U(q)\pi(dq)\\
&= -2\gamma\sum_i \alpha_i^2\int_{\mathbb{R}^n} q_i\partial_{q_i}\pi(q)dq - 2\gamma\sum_{i\neq j}\alpha_i\alpha_j\int_{\mathbb{R}^n} q_i\partial_{q_j}\pi(q)dq\\
&= 2\gamma\sum_i \alpha_i^2\int_{\mathbb{R}^n}\pi(q)dq - 2\gamma\sum_{i\neq j}\alpha_i\alpha_j\int_{\mathbb{R}^{n-1}} q_i\int_{\mathbb{R}}\partial_{q_j}\pi(q)dq_j dq_{-j}\\
&= 2\gamma\sum_i \alpha_i^2.
\end{aligned}$$

where dq_{-j} denotes $dq_1\dots dq_{j-1}dq_{j+1}\dots dq_n$. Taking $\gamma\to 0$ gives (3.4.15). ☐

Remark 3.6.1. In the proof of Proposition 3.4.7, either of

$$\gamma(q) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(q+\alpha)^2} \quad \text{or} \quad \gamma(q) = \frac{1}{\alpha\sqrt{2\pi}}e^{-\frac{1}{2}q^2}$$

for $\Gamma = \gamma I_n$ also work for large $\alpha > 0$ in place of $\gamma(q) = \epsilon$.

For the proof of Proposition 3.4.9, some notation is introduced. For $\tilde{k} \in \mathbb{N}_0$, let the tridiagonal matrix $M_{\tilde{k}} \in \mathbb{R}^{(\tilde{k}+1)\times(\tilde{k}+1)}$ be given by its elements

$$(M_{\tilde{k}})_{i,j} = \begin{cases} i & \text{if } i+1 = j, \\ (i-1)\gamma & \text{if } i = j, \\ i - \tilde{k} - 2 & \text{if } i-1 = j, \\ 0 & \text{otherwise} \end{cases} \tag{3.6.21}$$

for indices $1 \leq i, j \leq \tilde{k} + 1$.

Lemma 3.6.4. *Let $m \in \mathbb{N}$. Any tridiagonal matrix $\tilde{M} \in \mathbb{R}^{m\times m}$ of the form*

$$(\tilde{M})_{i,j} = \begin{cases} b_i & \textit{if } i+1 = j, \\ b_i'\gamma & \textit{if } i = j, \\ b_i'' & \textit{if } i-1 = j, \\ 0 & \textit{otherwise} \end{cases}$$

for constants $b_i, b_i', b_i'' \in \mathbb{R}$, has an order γ determinant as $\gamma \to 0$ if m is odd and a determinant that is bounded away from zero as $\gamma \to 0$ if m is even.

Lemma 3.6.4 is straightforwardly proved by repeatedly taking Laplace expansions. An explicit proof is not given here.

Proof of Proposition 3.4.9. Only a standard Gaussian and $M = 1$ is considered, the arguments for the general centered Gaussian case are the same. First consider the observable

$$f(q) = q^k \tag{3.6.22}$$

for some odd $k \in \mathbb{N}_0$. Take the polynomial ansatz

$$\phi(q,p) = \sum_{i,j=0}^{k} a_{i,j} q^i p^j \tag{3.6.23}$$

for $a_{i,j} \in \mathbb{R}$ and $\Gamma = \gamma > 0$. It will be shown that arbitrarily small asymptotic variance is achieved in the $\gamma \to 0$ limit. Note that only pairs (i,j) with odd i and even j make nonzero contributions to the asymptotic variance. Applying $-\mathcal{L}$ to the ansatz,

$$
\begin{aligned}
-\mathcal{L}\phi &= \sum_{i,j=0}^{k} -ia_{i,j}q^{i-1}p^{j+1} + ja_{i,j}q^{i+1}p^{j-1} + \gamma j a_{i,j}q^ip^j - \gamma j(j-1)a_{i,j}q^ip^{j-2} \\
&= \sum_{i,j=0} (-(i+1)a_{i+1,j-1} + (j+1)a_{i-1,j+1} + \gamma j a_{i,j} \\
&\qquad - \gamma(j+2)(j+1)a_{i,j+2})q^ip^j.
\end{aligned}
$$

where

$$a_{i,j} = 0 \quad \forall i,j < 0 \text{ and } \forall i,j > k. \tag{3.6.24}$$

Comparing coefficients in (3.1.5),

$$-(i+1)a_{i+1,j-1} + (j+1)a_{i-1,j+1} + \gamma j a_{i,j} - \gamma(j+2)(j+1)a_{i,j+2} = 0 \tag{3.6.25}$$

for all $(i,j) \neq (k,0)$. It holds by strong induction (in j') that

$$a_{i'+j',k+1-j'} = 0 \quad \forall i',j' \geq 0 \tag{3.6.26}$$

because of the following. The base case $j' = 0$ follows by (3.6.24), the induction step follows by taking $(i,j) = (i'+j'-1, k+2-j')$ for $i' \geq 0$ in (3.6.25) and again using (3.6.24) where necessary. Comparing coefficients in the Poisson equation (3.1.5) for $(i,j) = (k,0)$ and using (3.6.24), (3.6.26) yields[4]

$$a_{k-1,1} = 1. \tag{3.6.27}$$

Combining (3.6.27) with setting $(i,j) = (j'-1, k+1-j')$ for $j' = 1,\dots,k$ in (3.6.25), the entries $a_{j',k-j'}$ satisfy the linear system

$$M_k(a_{k,0}, a_{k-1,1}, \dots, a_{0,k})^\top = (1,0,\dots,0)^\top, \tag{3.6.28}$$

where $M_k \in \mathbb{R}^{k+1\times k+1}$ is the tridiagonal matrix given in (3.6.21). In order to find the order in γ as $\gamma \to 0$ of the elements of $(a_{k,0},\dots,a_{0,k})^\top$ appearing in (3.6.28), it suffices to find the order of the entries in the leftmost column of M_k^{-1}. For this, let $C_i \in \mathbb{R}$ be the i^{th} minor appearing in the top row of the cofactor matrix of M_k. On the corresponding submatrix, repeatedly taking the Laplace expansion on the leftmost column until only the determinant of a $(k+1-i)$-by-$(k+1-i)$ square matrix from the bottom right corner of M_k remains to be calculated, then using Lemma 3.6.4 for this $(k+1-i)$-by-

[4] It is illustrative to imagine a grid of coefficients and the relations (3.6.25) and (3.6.27) as L-shaped chains on the grid, where (3.6.26) and (3.6.24) leave only a triangular area of nonzero coefficients.

$(k+1-i)$ matrix gives that C_i is of order γ as $\gamma \to 0$ for odd i. Furthermore, the determinant of M_k is bounded away from zero as $\gamma \to 0$ by Lemma 3.6.4. Therefore the elements of $(a_{k,0},\ldots,a_{0,k})$ in the left hand side of (3.6.28) with an odd index, that is $a_{k-j,j}$ for even j, have order γ and at most order 1 otherwise as $\gamma \to 0$. These elements with odd indices are exactly those from the vector $(a_{k,0},\ldots,a_{0,k})^\top$ that make a contribution to the asymptotic variance. The 'next' set of contributions come from the vector $(a_{k-2,0},a_{k-3,1}\ldots,a_{0,k-2})$. Using again (3.6.24) and (3.6.25), the vector satisfies

$$M_{k-2}(a_{k-2,0},a_{k-3,1},\ldots,a_{0,k-2})^\top = v_{k-2},$$

for some vector v_{k-2} (from the last term on the left hand side of (3.6.25)) of order γ as $\gamma \to 0$ and since the determinant of M_{k-2} is of order 1 (by Lemma 3.6.4), the contributions here to the asymptotic variance are again of order γ. Continuing for

$$(a_{k-2j,0},a_{k-2j-1,1}\ldots,a_{0,k-2j})^\top, \quad j \in \mathbb{N},$$

it follows that all contributions are of order γ as $\gamma \to 0$. The resulting coefficients indeed make up a solution ϕ to the Poisson equation because the matrices M_k are invertible and because the coefficients $a_{i,j}$ for even $i+j$ are equal to zero from repeating the above procedure for the coefficients associated to M_{k-1}, M_{k-3} and so on.

For the general case of (3.4.16), since $\mathcal{L}$ is a linear differential operator and the contributions to the value of $\int \phi(f-\pi(f))d\tilde{\pi}$ come from exactly the same (odd i, even j) $a_{i,j}$ coefficients from the corresponding solution ϕ to each summand in (3.4.16), the proof concludes. $\square$

Proof of Proposition 3.4.10. Take the polynomial ansatz

$$\phi(q,p) = \sum_{i,j=0}^{4} a_{i,j} q^i p^j \tag{3.6.29}$$

for $a_{i,j} \in \mathbb{R}$, where $a_{i,j}$ not appearing in the sum are taken to be zero in the following. Again, only the standard Gaussian is considered, it turns out the arguments follow similarly otherwise. Comparing coefficients in (3.1.5) and using the same strong induction argument as in the proof of Proposition 3.4.9 leads to (3.6.25) for all $(i,j) \neq (4,0),(0,0)$ and equation (3.6.26). Taking $(i,j) = (j'-1, 5-j')$ for $1 \leq j' \leq 4$ in (3.6.25) and comparing the q^4 coefficients in the Poisson equation, it holds that

$$M_4(a_{4,0},a_{3,1},a_{2,2},a_{1,3},a_{0,4})^\top = (1,0,\ldots,0)^\top \tag{3.6.30}$$

and taking $(i,j) = (j'-1, 3-j')$ for $j' \geq 1$ in (3.6.25) yields

$$M_2(a_{2,0}, a_{1,1}, a_{0,2})^\top = \gamma(2a_{2,2}, 6a_{1,3}, 12a_{0,4})^\top. \tag{3.6.31}$$

Equations (3.6.30), (3.6.31) can be solved explicitly and the asymptotic variance is a weighted sum of the resulting coefficients. Those in (3.6.29) that make contributions are $a_{4,0}, a_{2,2}, a_{2,0}$, which gives the asymptotic variance $\frac{12(21\gamma^4+55\gamma^2+27)}{\gamma(3\gamma^2+4)}$ that goes to infinity as $\gamma \to 0$ or $\gamma \to \infty$. Comparing constant terms in the Poisson equation yields

$$a_{0,2} = \frac{1}{2\Gamma} \int q^4 \frac{e^{\frac{q^2}{2}}}{\sqrt{2\pi}} dq = \frac{3}{2\Gamma},$$

which turns out to be satisfied by the solution for $a_{0,2}$, so that (3.6.29) is indeed a solution; note that the coefficients associated to M_3 and M_1 are zero by a similar procedure as above. $\square$

Proof of Proposition 3.4.11. Note for a quadratic ϕ, it can be read immediately from condition (3.4.7) that the nonconstant part of Γ must be equal to zero. Therefore consider

$$\phi(q,p) = \sum_{i,j=0} a_{i,j} q^i p^j \tag{3.6.32}$$

where $a_{i,j} \in \mathbb{R}$ for all $i, j \geq 0$. The function Γ is given by

$$\Gamma(q) = \sum_{k \in \mathbb{N}_0} b_k q^k > 0 \tag{3.6.33}$$

for $b_k \in \mathbb{R}$, $k \in \mathbb{N}$, $b_0 > 0$ and

$$\begin{aligned} -\mathcal{L}\phi = -\sum_{i,j=0} \Bigg(&(i+1)a_{i+1,j-1} - (j+1)\Sigma^{-1} a_{i-1,j+1} - j \sum_{k=0} a_{i-k,j} b_k \\ &+ (j+2)(j+1) \sum_{k=0} a_{i-k,j+2} b_k \Bigg) q^i p^j, \end{aligned}$$

where if either $i < 0$ or $j < 0$, we set $a_{i,j} = 0$. Comparing coefficients in q, p in the

Poisson equation (3.1.5), we obtain

$$2a_{0,2}b_0 = \frac{1}{2}U_0\Sigma \tag{3.6.34}$$

$$\Sigma^{-1}a_{0,1} - 2\sum_{k=0}^{1} a_{1-k,2}b_k = l \tag{3.6.35}$$

$$\Sigma^{-1}a_{1,1} - 2\sum_{k=0}^{2} a_{2-k,2}b_k = \frac{1}{2}U_0 \tag{3.6.36}$$

$$\begin{aligned}(i+1)a_{i+1,j-1} - (j+1)\Sigma^{-1}a_{i-1,j+1} - j\sum_{k=0}^{i} a_{i-k,j}b_k \\ +(j+2)(j+1)\sum_{k=0}^{i} a_{i-k,j+2}b_k = 0 \qquad \forall (i,j) \neq (0,0),(1,0),(2,0).\end{aligned} \tag{3.6.37}$$

Suppose for contradiction there exists a finite order polynomial solution to (3.1.5). Let

$$j^* := \min\{J \in \mathbb{N}_0 : a_{i,j} = 0 \quad \forall j \geq J, i \geq 0\} \tag{3.6.38}$$

and $k^* \in \mathbb{N}_0$ be the order of the polynomial Γ, that is,

$$b_k = 0 \quad \forall k \geq k^* + 1, \qquad b_{k^*} \neq 0. \tag{3.6.39}$$

Note that $k^* \geq 2$ since Γ must be positive. First of all, ϕ clearly cannot be a function of just q. If $j^* = 2$, then condition (3.6.37) applied to $i = 0$, $j = 2$ gives $a_{1,1} = 0$ which contradicts with equation (3.6.36). Therefore $j^* \geq 3$. In this case, condition (3.6.37) applied to $j = j^*$ yields

$$a_{i+1,j^*-1} = 0 \quad \forall i \geq 0, \tag{3.6.40}$$

which in turn by condition (3.6.37) this time applied to $j = j^* - 1$ yields

$$a_{i+1,j^*-2} = 0 \quad \forall i \geq k^* + 1.$$

A strong induction argument yields for all $1 \leq j \leq j^* - 1$ the equation

$$a_{i+1,j-1} = 0 \quad \forall i \geq (j^* - j)(k^* + 1),$$

or equivalently, using the definition (3.6.38) for j^*, for all $j \geq 1$ the equation

$$a_{i+1,j-1} = 0 \quad \forall i \geq \max\{-1, (j^* - j)(k^* + 1)\} \tag{3.6.41}$$

and in particular

$$a_{i+1,2} = 0 \quad \forall i \geq (j^* - 3)(k^* + 1).$$

Now with this and (3.6.39), applying condition (3.6.37) for $(i, j) = ((j^* - 2)(k^* + 1) + 1, 0)$ gives

$$a_{(j^*-2)(k^*+1),1} = 0,$$

but then condition (3.6.37) for $(i, j) = ((j^* - 2)(k^* + 1) - 1, 2)$, together with (3.6.41) for $j = 2, 3, 4$ gives

$$a_{(j^*-3)(k^*+1),2} = 0.$$

If $j^* = 3$, then this contradicts already since $a_{i,2} = 0$ for all $i \geq 0$ and so $j^* \leq 2$ by definition. Carrying on from this for $j^* > 3$, in a similar way, condition (3.6.37) for $(i, j) = ((j^* - 3)(k^* + 1) - 1, 3)$, with (3.6.41) for $j = 3, 4, 5$, gives

$$a_{(j^*-4)(k^*+1),3} = 0$$

and so on, where the last equation from the induction, which use condition (3.6.37) for $i = k^*$, $j = j^* - 1$ and (3.6.41) for $j = j^* - 1, j^*, j^* + 1$, is

$$a_{0,j^*-1} = 0.$$

Together with (3.6.40), this contradicts with the definition (3.6.38) of j^*. $\square$

3.7 Discussion

3.7.1 Relation to literature

The infinite time integral (3.1.9) has been used for the calculation of transport coefficients in molecular dynamics [120, 157] and the derivative of the expectation appearing in (3.1.9) with respect to initial conditions is a problem considered when calculating the 'greeks' in mathematical finance [66]. On the topic of the latter and in contrast to [66], there is previous work dealing with cases of degenerate noise in the system, but the formulae derived were done so under different motivations and do not seem to improve upon (3.1.10) in our situation; some of these references are given in Remark 3.5.2.

Taking $\Gamma \to \infty$ together with a time rescaling, the dynamics (3.1.1) become the overdamped Langevin equation [158]. An analogous result holds [98] when $\Gamma = \Gamma(q)$ is position dependent, where a preconditioner for the corresponding overdamped dynamics appears in terms of Γ^{-1}; see Section 3.7.3 for a consideration of our method in the position dependent friction case. On the other hand, the Hessian of U makes a good preconditioner in the overdamped dynamics because of the Brascamp-Lieb inequality, see Remark 1 in [3].

On the application of underdamped Langevin dynamics with (variance reduced) stochastic gradients alongside the related Hamiltonian Monte Carlo method, [197] presents a comparison with convergence rates for the latter. In [39], convergence guarantees are provided for variance reduced gradients in the overdamped case and the control variate stochastic gradients in the underdamped case, along with numerical comparisons in low dimensional, tall dataset regimes. Furthermore, the underdamped dynamics with single, randomly selected component gradient update in place of the full gradient is considered in [50].

Variance reduction by modifying the observable instead of changing the dynamics has been considered for example in [7, 9, 180]. The methods there are incompatible with the framework in the present work due to the improved observable being unknown before the simulation of the Markov chain. Although useful, their applicability are limited in large n cases due to storage requirements [7], not to mention either escalating computational cost for improvements in the observable or requirement of a priori knowledge [180].

On the topic of adaptive MCMC, we refer the reader to the review in [79].

3.7.2 The nonconvex case

In the case where U is nonconvex, the Monte Carlo procedure in Section 3.5.1 may continue to be used as presented, however the first variation process could easily stray from the case of exponential decay as in Theorem 3.3.3. Transitions from one metastable state to another cause the tangent process to increase in magnitude. In a one dimension double well potential $U(q) = \frac{q^4}{4} - q^2 + \frac{q}{2}$, linear observable $f(q) = q$ case, these transitions occur frequently enough during the gradient procedure in Γ that Dq blows up in simulation. Even in cases for which the metastabilities are strong, so that transitions occur less frequently, simulations show that Γ dives to zero in periods where no transitions are occuring (as if the case of Corollary 3.4.8), but increase dramatically in value once a transition does occur, causing the trajectory in Γ to decay over time but occasionally jumping in value, so that there is no convergence for Γ. On the other hand, the Galerkin method presented in Section 3.5.5 tends to give good convergence for Γ in such cases.

3.7.3 Position-dependent friction

It is possible to adapt the formula (3.3.1) to the case of position-dependent gradient direction in Γ. The gradient direction is the same as (3.1.8) with the change that the integral is replaced by the corresponding marginal integral in p. Ideas using such a formula need to take into account that the first variation process retains a non-vanishing stochastic integral with respect to Brownian motion, so that the truncation in calculating the corresponding infinite time integral in Section 3.5.1 is not as well justified, or rather, does not happen in the execution of Algorithm 2 due to (3.5.11) not being satisfied. Moreover, the work here raises the same question under the setting where the friction matrix depends on both position and velocity, which is the original motivating factor for the next chapter.

3.7.4 Metropolisation

Throughout Section 3.5, the implementation has not involved accept-reject steps. Metropolisation of discretisations of the underdamped Langevin dynamics was given in [97], see also Section 2.2.3.2 in [123] and [141, 174]. The systematic discretisation error is removed with the inclusion of this step but the momentum is reversed upon rejection (to avoid high rejection rates [174]), which raises the question of whether friction matrices arising from Algorithm 1 improve the Metropolised situation where dynamics no longer imitate those in the continuous-time. For example the intuition in the Gaussian target measure, linear observable case discussed in Section 3.4.2 no longer applies.

3.7.5 Conclusion

We have presented the central limit theorem for the underdamped Langevin dynamics and provided a formula for the directional derivative of the corresponding asymptotic variance with respect to a friction matrix Γ. A number of methods for approximating the gradient direction in Γ have been discussed together with numerical results giving improved observed variances. Some cases where an improved friction matrix can be explicitly found have been given to guide the expectation of an optimal Γ. In particular, in cases where the observable is linear and the potential is close to quadratic, which is the case when finding the posterior mean in Bayesian inference with Gaussian priors, the optimal friction is expected to be close to zero (due to Corollary 3.4.8). This is consistent with the numerical conclusion from the proposed Algorithm 2. Moreover, it is shown that the improvement in variance is retained when using minibatch stochastic gradients in a case of Bayesian inference.
We mention that the gradient procedure using (3.1.6) and (3.1.10) can be used to guide Γ in arbitrarily high dimension by extrapolation; that is, given a high dimensional problem of interest, the gradient procedure can be used on similar, intermediate dimensional problems in order to obtain a friction matrix that can be extrapolated to the original problem. In particular, for the Bayesian inference problem as formulated in Section 3.5.4, the algorithm recommends the choice of a small friction scalar, which can be expected to

apply for datasets in an arbitrary number of dimensions.

Future directions not mentioned above includes well-posedness of the optimisation in Γ, extension to higher-order Langevin samplers methods as in the previous chapter or [144] and gradient formulae in the discrete time case analogous to Theorem 3.3.2.

4

Regularity preservation in Kolmogorov equations under Lyapunov conditions

The contents of this chapter are from the paper [35].

4.1 Introduction

Consider for $b : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and a standard Wiener process W_t, the SDE on $\mathbb{R}^n$ given by

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{4.1.1}$$

The underlying results of this work are moment bounds of derivatives of X_t with respect to initial condition in the case that the coefficients b and σ are not globally Lipschitz continuous in space. These estimates are used to validate an Itô-Alekseev-Gröbner formula [99] and differentiability of semigroups associated with (4.1.1), which enable the existence of twice differentiable-in-space solutions to Kolmogorov equations [114], solutions to corresponding Poisson equations, weak error estimates of numerical approximations [112, 185, 192] and related derivative estimates [46]. Similar moment bounds on the first and second derivative with respect to initial value in the non-globally Lipschitz case have recently been studied in [100] using the stochastic Grönwall inequality [101, 175], where related ideas for the non-globally monotone case had appeared earlier in [44, 103]. In this work, it is shown that the above consequences hold for the case that the coefficients b, σ are non-globally monotone, where higher derivatives of b, σ are bounded by Lyapunov functions and loosely that b and σ admit Lipschitz constants which are $o(\log V)$ and $o(\sqrt{\log V})$ respectively for a Lyapunov function V. The results are applicable to all of the example SDEs in [105] except in Section 4.7, in Section 4.7.2 we consider specifically the stochastic Duffing-van der Pol equation given by

$$dx_t = y_t dt \tag{4.1.2a}$$

$$dy_t = [\alpha_1 x_t - \alpha_2 y_t - \alpha_3 y_t x_t^2 - x_t^3]dt + \beta_1 x_t dW_t + \beta_3 d\tilde{W}_t, \tag{4.1.2b}$$

where $\alpha_1, \alpha_2, \beta_1, \beta_3 \in \mathbb{R}$, $\alpha_3 > 0$ and W_t, $\tilde{W}_t$ are independent standard Wiener processes. In particular, for the first time, weak convergence rates of order one are shown for a numerical scheme approximating SDEs with non-globally monotone coefficients including the stochastic Duffing-van der Pol oscillator as given above, stochastic Lorenz equation with additive noise, underdamped Langevin equation with variable friction, overdamped Langevin dynamics with a non-globally monotone potential gradient and the stochastic Ginzburg-Landau equation.

In contribution to regularity analysis of SDEs, the results give criteria for the positive case of regularity for semigroups beyond the settings of globally Lipschitz coefficients or ellipticity and against the counterexamples of [85] with globally bounded smooth coefficients. For example, one of the counterexamples given in the aforementioned reference is the SDE

$$\begin{aligned} dx_t &= \cos(z_t e^{y_t^3})dt \\ dy_t &= \sqrt{2}dW_t \\ dz_t &= 0dt, \end{aligned}$$

which turns out to have the property that $\mathbb{E}[\varphi(x_t, y_t, z_t)]$ is not locally Hölder continuous in initial value for some compactly supported smooth φ. In the present work, the coefficients b and σ are not directly imposed to be in any weighted L^∞ spaces for spatial differentiability of semigroups, note however the coefficients are indirectly bounded by the aforementioned local Lipschitz bound. More generally, the regularity with respect to initial condition demonstrated here under our assumptions has further counterexamples for SDEs with constant diffusion coefficient and drift with polynomially growing first derivatives [109]. On the other hand, in the globally Lipschitz case [114] and the globally monotone (or one-sided Lipschitz) case [34], moment bounds on derivatives with respect to initial condition are known. In addition, for infinitely differentiable b and σ satisfying Hörmander's bracket condition [95], infinite differentiability of the associated semigroup is given by Proposition 4.18 in [85].

Our basic result about semigroup differentiability can be summarised as follows.

Assumption 6. There exists $V : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable in space, continuously differentiable in time and constant $C > 0$ such that $\partial_t V(t,x) +$

$\sum_{i=1}^n b_i(t,x)\partial_{x_i}V(t,x)+\frac{1}{2}\sum_{i,j=1}^n \sigma(t,x)\sigma(t,x)^\top\partial_{x_i}\partial_{x_j}V(t,x) \leq CV(t,x)$ for all $t\in[0,T]$, $x\in\mathbb{R}^n$ and $\lim_{|x|\to\infty}V(t,x)=\infty$. The functions $f,c:[0,T]\times\mathbb{R}^n\to\mathbb{R}$, $g:\mathbb{R}^n\to\mathbb{R}$ are measurable functions and $p\in\mathbb{N}_0$. For any $R>0$, it holds that $\int_0^T\sup_{x\in B_R}(|c(t,x)|+|f(t,x)|)dt<\infty$, $b,\sigma,f(t,\cdot),g,c(t,\cdot)\in C^p$, and

- there exists measurable $G:[0,T]\times\mathbb{R}^n$ such that $G(t,\cdot)=o(\log V(t,\cdot))$ uniformly in t and satisfying

$$|b(t,x)-b(t,y)|\leq (G(t,x)+G(t,y))|x-y|,$$
$$\|\sigma(t,x)-\sigma(t,y)\|^2\leq (G(t,x)+G(t,y))|x-y|^2$$

  for all $t\in[0,T]$, $x\in\mathbb{R}^n$,

- for every $k>0$, $h\in\{b,f,g,c\}$, there exists $C'>0$ such that

$$|\partial^\alpha h(t,\lambda x+(1-\lambda)y)|+\|\partial^\beta\sigma(t,\lambda x+(1-\lambda)y)\|^2\leq C'(1+V(t,x)+V(t,y))^{\frac{1}{k}}$$

  for all $t\in[0,T]$, $x,y\in\mathbb{R}^n$, $\lambda\in[0,1]$ and multiindices α,β with $p_0\leq|\alpha|\leq p$, $2\leq|\beta|\leq p$, where $p_0=2$ if $h=b$ and $p_0=0$ otherwise.

To comment on Assumption 6, the first assertion about the existence of G is our main assumption about the Lipschitz constant of the coefficients of the SDE and generalises the global Lipschitz condition; the second assertion about the higher derivatives is less stringent in term of the right-hand side, but more so in the sense that the bound should hold in between two points x,y as λ varies.

Theorem 4.1.1. *Let Assumption 6 hold. For any $s\in[0,T]$ and stopping time $\tau\leq T-s$, the expectation of the random function $u(s,\tau,\cdot):\mathbb{R}^n\to\mathbb{R}$ given by*

$$\mathbb{E}u(s,\tau,x)=\mathbb{E}\bigg[\int_0^\tau f(s+r,X_r^{s,x})e^{-\int_0^r c(s+w,X_w^{s,x})dw}dr+g(X_\tau^{s,x})e^{-\int_0^\tau c(s+w,X_w^{s,x})dw}\bigg], \tag{4.1.3}$$

is continuously differentiable in x up to order p, where for any $s\in[0,T]$, $x\in\mathbb{R}^n$, $X_\cdot^{s,x}$ is the solution to $X_t^{s,x}=x+\int_0^t b(s+r,X_v^{s,x})dr+\int_0^t\sigma(s+r,X_r^{s,x})dW_r$ on $t\in[0,T-s]$. Moreover, if $p\geq 2$, the function given by $v(t,x)=\mathbb{E}u(t,T-t,x)$ is locally Lipschitz in t and satisfies

$$\partial_t v+a:D^2v+b\cdot\nabla v-cv+f=0 \tag{4.1.4}$$

almost everywhere in $(0,T)\times\mathbb{R}^n$.

Results assuming instead local estimates on the derivatives of b,σ,f,c,g in the case of time-independent b,σ are given in Section 4.5. Theorem 4.1.1 follows as corollary to Theorem 4.4.2 and the assertion about Kolmogorov equations can be found in Theorem 4.4.5.

The latter appears to be the only result against the literature about almost everywhere twice differentiable-in-space solutions to Kolmogorov equations in the non-hypoelliptic, non-elliptic diffusion coefficient and non-globally Lipschitz coefficients case.
Our basic result about weak convergence rates of a numerical approximation is as follows, assuming a deterministic initial condition for (4.1.1). The scheme considered here is the stopped increment-tamed Euler-Maruyama scheme from [105] and it is given by

$$Y_t^\delta = Y_{k\delta}^\delta + \mathbb{1}_{\{|Y_{k\delta}^\delta| \leq \exp(|\log\delta|^{\frac{1}{2}})\}} \left(\frac{b(Y_{k\delta}^\delta)(t-k\delta) + \sigma(Y_{k\delta}^\delta)(W_t - W_{k\delta})}{1 + |b(Y_{k\delta}^\delta)(t-k\delta) + \sigma(Y_{k\delta}^\delta)(W_t - W_{k\delta})|^4} \right). \tag{4.1.5}$$

Theorem 4.1.2. *Let Assumption 6 hold with $p \geq 3$. Suppose b, σ are independent of t and suppose V is of the form $V(t,x) = e^{U(x)e^{-\rho t}}$ for $U \in C^3(\mathbb{R}^n, [0,\infty))$, $\rho > 0$, such that there exist $c \geq 1$ satisfying*

$$|x|^{\frac{1}{c}} + |\partial^\alpha b(x)|^{\frac{1}{c}} + \|\partial^\alpha \sigma(x)\|^{\frac{1}{c}} + |\partial^\beta U(x)| \leq c(1+U(x))^{1-\frac{1}{c}},$$

for all $x \in \mathbb{R}^n$, multiindices α, β with $0 \leq |\alpha| \leq 2$ and $1 \leq |\beta| \leq 3$. If $h \in C^3(\mathbb{R}^n, \mathbb{R})$ is such that

$$|\partial^\alpha h(x)| \leq c(1+|x|^c)$$

for all $x \in \mathbb{R}^n$ and multiindices α with $0 \leq |\alpha| \leq 3$, then there exists a constant $C > 0$ such that

$$|\mathbb{E}[h(X_T)] - \mathbb{E}[h(Y_T^\delta)]| \leq C\delta, \tag{4.1.6}$$

for all $0 < \delta < 1$, where $Y_\cdot^\delta : [0,T] \to \mathbb{R}^n$ is the approximation given by $Y_0^\delta = X_0$ and (4.1.5) for all $t \in [k\delta, (k+1)\delta]$, $k \in \mathbb{N}_0 \cap [0, \frac{T}{\delta})$.

Theorem 4.1.2 is corollary to Theorem 4.6.3. The numerical scheme (4.1.5) from [105] has the key property of retaining exponential integrability properties of the continuous time SDE, which is used throughout the proof for Theorem 4.6.3. As is well documented [104], the classical Euler-Maruyama scheme may diverge in both the strong and weak sense for superlinearly growing, non-globally Lipschitz coefficients without this property. The proof of Theorem 4.6.3 uses the recent Itô-Alekseev-Gröbner formula [99] in order to expand the left-hand side of (4.1.6), instead of using a solution to Kolmogorov equations as in [112] that is twice continuously differentiable-in-space and once continuously differentiable-in-time. The Itô-Alekseev-Gröbner formula describes in particular the pathwise error associated to the left-hand side of (4.1.6) in terms of local differences in the coefficients of the respective SDEs for X_t and Y_t^δ, of the derivative processes (of X_t) and of derivatives of h. In order to satisfy the assumptions for the formula, strong completeness using a result in [44] is shown for the derivative processes; note this property appeared recently

under the results of [100] using a different approach and different assumptions. Although weak convergence without rates has been established by way of convergence in probability in [105, Corollary 3.7] and [102, Corollary 3.19], weak rates of convergence analogous to the globally monotone case (of order 1 as above) seem to be an open problem for non-globally monotone coefficients outside of the present work. For weak convergence results in the former case, see references within [44] and also [192]. On the other hand, strong convergence rates of order $\frac{1}{2}$ have been established in even the latter case, see [103].
The proofs for the moment estimates underlying both Theorems 4.1.1 and 4.1.2 use directly the results of [101], where the strongest assumptions that are made here are used for exponential integrability as in [44, 103]. The core argument is to consider for any $\kappa \in \mathbb{R}^n$ processes $X_{t(\kappa)}$ satisfying

$$\sup_{t\in[0,T]}\left|\frac{X_t^{x+r\kappa} - X_t^x}{r} - X_{t(\kappa)}^x\right| \to 0$$

in probability as $r \to 0$, where X_t^x denotes a solution to (4.1.1) with $X_0^x = x$. Such processes exist [114, Theorem 4.10] for b, σ continuously differentiable in space satisfying some local integrability assumption and $X_{t(\kappa)}^x$ satisfies the system resulting from a formal differentiation of (4.1.1) (see precisely (4.3.5)). If b and σ are independent of t and the derivatives of b and σ are locally Lipschitz, the processes $X_{t(\kappa)}^x$ are almost surely continuous derivatives in the classical sense as in [165, Theorem V.39]. Higher derivatives exist for b and σ with higher orders of differentiability. Considering SDEs solved by these derivatives, it is seen that only the term involving the derivative of the highest order on the right-hand side of the dynamics requires serious control and that the stochastic Grönwall inequality of [101] together with our Lipschitz Assumption 7 and an induction argument suffice to control all of the terms. The bounds for higher derivatives, as required for twice differentiability of the semigroup in the Kolmogorov equation for example, call for two-sided Lipschitz conditions as in our Assumption 7 in contrast to the one-sided conditions in [44, 103]. We use $o(\log V)$ and $o(\sqrt{\log V})$ Lipschitz constants in order to control the moments for large time T, but the results follow for $O(\log V)$ and $O(\log V)$ Lipschitz constants if T is suitably small. In order to establish solutions to the Kolmogorov equation for the setting here, we prove a number of intermediary results following the strategy of [114]. In particular, we show by extending an argument in [164] that an Euler-type approximation converges to solutions of the SDE in probability and locally uniformly in initial time and space, that is, the SDE is regular [114, Definition 2.1], which is also used as mentioned for Theorem 4.1.2.
In relation to the previous chapter, our result about the Kolmogorov equation is motivated by its use in establishing that the semigroup associated to Langevin dynamics

indeed forms a solution to the Poisson equation in the distributional sense, even when maximal dissipativity results on the generator are not available and beyond the globally Lipschitz case. In particular, this is the case for the Langevin equation with position and velocity-dependent friction. Indeed, we have that given an invariant measure $\tilde{\pi}$, $f \in L^2(\tilde{\pi})$, $h \in C_c^\infty(\mathbb{R}^{2n})$, generator $\mathcal{L}$ defined for twice differentiable functions, its $L^2(\tilde{\pi})$-adjoint $\mathcal{L}^*$, $\phi(t,x) = \int_0^\infty \mathbb{E} f(X_t^x)dt$, $P_t f(x) = \mathbb{E} f(X_t^x)$, and an approximating sequence $(f_k)_{k\in\mathbb{N}}$ such that $f_k \in C_c^\infty(\mathbb{R}^{2n})$, $f_k \to f$ in $L^2(\tilde{\pi})$, it holds that

$$\begin{aligned}
\left|\int \mathcal{L}^* h\phi d\tilde{\pi} - \int hfd\tilde{\pi}\right| \leq& \left|\int \mathcal{L}^* h\phi d\tilde{\pi} - \int \mathcal{L}^* h \int_{\frac{1}{T}}^T P_t(f)dtd\tilde{\pi}\right| \\
&+ \left|\int h(f - P_{\frac{1}{T}}(f) + P_T(f))d\tilde{\pi}\right| \\
&+ \left|\int \mathcal{L}^* h \int_{\frac{1}{T}}^T P_t(f-f_k)dtd\tilde{\pi}\right| \\
&+ \left|\int h(P_{\frac{1}{T}}(f-f_k) - P_T(f-f_k))d\tilde{\pi}\right| \\
&+ \left|\int \mathcal{L}^* h \int_{\frac{1}{T}}^T P_t(f_k)dtd\tilde{\pi} + \int h(P_{\frac{1}{T}}(f_k) - P_T(f_k))d\tilde{\pi}\right|, \quad (4.1.7)
\end{aligned}$$

so that, given convergences of the first terms, the semigroup $P_t f$ being a solution of the Kolmogorov equation is enough to conclude that the Poisson equation is solved by ϕ in the distributional sense.

Alternative to the related Feynman-Kac formula for making the connection between the SDE and the Kolmogorov equation satisfied by the transition semigroup is to use the theory of Dirichlet forms. In that context, the order in which one works is reversed, in the sense that one begins with the partial differential operator acting on the space of smooth compactly supported functions, shows that the closure of the graph in some Banach space generates a strongly continuous semigroup and that the semigroup is associated to the transition semigroup of a Hunt process that is a weak solution to the stochastic differential equation. In the case of Langevin dynamics in [11, 43], note however that the Banach space is $L^2(\pi)$ for the invariant measure π, so that the strongly continuous semigroup is a classical solution in the sense of an abstract Cauchy problem, where the transition semigroup is then implied to be in the domain of the infinitesimal generator, but not necessarily a twice differentiable-in-space function solving the partial differential equation pointwise nor almost everywhere nor in the sense of distributions.

For solutions in the sense of distributions to the Kolmogorov equation, in Proposition 4.18 in [85], the authors make use of (essentially) Lemma 5.12 in [114] to obtain such solutions in the case of smooth coefficients. In fact, the proof there only makes use of local

Lipschitz continuity of the coefficients and that $\sigma\sigma^\top$ admits a derivative which is locally Lipschitz. The authors in [85] moreover provide existence and uniqueness results for viscosity solutions under a Lyapunov condition, which yield distributional solutions under Hölder regularity on the coefficients [108] and almost everywhere solutions given enough regularity on the solution, see Proposition I.4 in [127]. On a similar note, the existence of a viscosity solution to the Poisson equation with possible degenerate second-order coefficients is given by [155], which, to the best of the author's knowledge, requires the aforementioned regularity to be interpreted as a solution in the distributional sense. In the present work, sufficient conditions for the backward Kolmogorov equation to hold in the distributional sense are presented as a secondary result. In particular, maximal dissipativity of the associated infinitesimal generator defined on smooth compactly supported functions is shown to be (mostly) sufficient.

The rest of the chapter is organised as follows. In Section 4.2, the setting, notation and various definitions about what is referred to as Lyapunov functions are given. In Section 4.3, moment estimates of the supremum over time on the derivative process and the difference processes in initial value are given. These results are used throughout for proving the other results in the chapter. In Section 4.4, results on the regularity of the semigroup associated to (4.1.1) are presented, which are followed by results about twice differentiable-in-space and distributional solutions to the Kolmogorov equation. Section 4.6 contains the results about weak convergence rates for the stopped increment-tamed Euler-Maruyama scheme on SDEs with non-globally monotone coefficients.

4.2 Notation and preliminaries

Just as the previous chapters, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, $\mathcal{F}_t$, $t \in [0,\infty)$, be a filtration satisfying the usual conditions and $(W_t)_{t\geq 0}$ be a standard Wiener process on $\mathbb{R}^n$ with respect to $\mathcal{F}_t$, $t \in [0,\infty)$. Let $T \in (0,\infty)$ and $\|M\|$ denote the Frobenius norm of a matrix M. Let $b : \Omega\times[0,\infty)\times\mathbb{R}^n \to \mathbb{R}^n$, $\sigma : \Omega\times[0,\infty)\times\mathbb{R}^n \to \mathbb{R}^{n\times n}$ be functions such that $b(t,\cdot), \sigma(t,\cdot)$ are continuous for every t,ω, $b(\cdot,x), \sigma(\cdot,x)$ are $\mathcal{F}\otimes\mathcal{B}([0,\infty))$-measurable for every x, $b(t,x), \sigma(t,x)$ are $\mathcal{F}_t$-measurable for every t,x and $\int_0^T \sup_{|x|\leq R}(|b(t,x)| + \|\sigma(t,x)\|^2)dt < \infty$ for any $R > 0$, $\omega \in \Omega$. For an open set $O \subseteq \mathbb{R}^n$ and any $x \in O$, let $X_t^{s,x}$ be an $\mathcal{F}_t$-adapted O-valued process such that $X_t^{s,x}$ is $\mathbb{P}$-a.s. continuous satisfying for all $s,t \in [0,T]$,

$$X_t^{s,x} = x + \int_0^t b(s+r, X_r^{s,x})dr + \int_0^t \sigma(s+r, X_r^{s,x})dW_r. \tag{4.2.1}$$

When the initial conditions are not important or are obvious from the context, simply X_t and similarly X_t^x is written. For $f \in C^2(O)$ and for either b, σ as above or $(b_\cdot^x :$

$\Omega \times [0,T] \to \mathbb{R}^n)_{x\in O}$, $(\sigma^x_\cdot : \Omega \times [0,T] \to \mathbb{R}^{n\times n})_{x\in O}$ that are, for each x, $\mathcal{F}\otimes\mathcal{B}([0,t])$-measurable and $\mathcal{F}_t$-adapted satisfying $\mathbb{P}$-a.s. that $\int_0^T(|b^x_s| + |\sigma^x_s|^2)ds < \infty$, we denote

$$Lf = b\cdot\nabla f + a : D^2 f, \tag{4.2.2}$$

where $a = \frac{1}{2}\sigma\sigma^\top$, D^2 denotes the Hessian and for matrices M, N, $M : N = \sum_{i,j} M_{ij}N_{ij}$. Throughout, $\hat{O}$ is used to denote the convex hull of O, $C^\infty_c((0,T)\times\mathbb{R}^n)$ denotes the set of compactly supported infinitely differentiable functions on $(0,T)\times\mathbb{R}^n$, $C_b(\mathbb{R}^n)$ denotes the set of bounded continuous function on $\mathbb{R}^n$, $C^{1,2}([0,T]\times\mathbb{R}^n)$ denotes the set of continuous functions of the form $[0,T]\times\mathbb{R}^n \ni (t,x)\mapsto f(t,x)$ that are once continuously differentiable in t and twice so in x, $B_R(x)$ denotes the closed ball of radius $R>0$ around $x\in\mathbb{R}^n$, $B_R = B_R(0)$, e_i denotes the i^{th} Euclidean basis vector in $\mathbb{R}^n$, and $C>0$ denotes a generic constant that may change from line to line. The expression $\mathbb{1}_A$ denotes the indicator function on the set A. We denote $\Delta_T = \{(s,t) : 0\le s\le t\le T\}$. The notation $\partial_i Z^z_{t,T} = \partial_{z_i} Z^\cdot_{t,T}|_z$ is used and similarly for the higher order derivatives $\partial^\alpha Z^z_{t,T}$ for multiindices α. Moreover, for a multiindex α, we denote

$$\kappa_\alpha = (e_1, \overbrace{\dots}^{\alpha_1 \text{ times}}, e_1, e_2, \dots).$$

Definition 4.2.1. A positive random function $V : \Omega\times[0,T]\times O \to (0,\infty)$ is referred to as a $(\tilde{b}^\cdot_\cdot, \tilde{\sigma}^\cdot_\cdot, \alpha_\cdot, \beta_\cdot, p^*, V_0)$-Lyapunov function if $\mathcal{F}\otimes\mathcal{B}([0,T])$-measurable and $\mathcal{F}_t$-adapted processes $\tilde{b}^\cdot_\cdot : \Omega\times[0,T]\times O\to\mathbb{R}^n$, $\tilde{\sigma}^\cdot_\cdot : \Omega\times[0,T]\times O \to \mathbb{R}^{n\times n}$, $\alpha_\cdot, \beta_\cdot : \Omega\times[0,T]\to[0,\infty]$, $p^*\in[1,\infty)$ and $V_0\in C^{1,2}([0,T]\times O)$ are such that for all $y\in O$ there exists an $\mathcal{F}\otimes\mathcal{B}([0,T])$-measurable, $\mathcal{F}_t$-adapted process $Y^y_\cdot : \Omega\times[0,T]\to O$ that is $\mathbb{P}$-a.s. continuous, $V(t,y) = V_0(t,Y^y_t)$ and it holds $\mathbb{P}$-a.s. that

$$\int_0^T (|\tilde{b}^y_s| + |\tilde{\sigma}^y_s|^2 + |\alpha_s|)ds < \infty,$$

$$Y^y_{t\wedge\tau} = y + \int_0^t \mathbb{1}_{[0,\tau)}(u)\tilde{b}^y_u du + \int_0^t \mathbb{1}_{[0,\tau)}(u)\tilde{\sigma}^y_u dW_u, \tag{4.2.3}$$

$$(\partial_t + L)V_0(t,Y^y_t) + \frac{p^*-1}{2}\frac{\left|(\tilde{\sigma}^y_t)^\top\nabla V_0(t,Y^y_t)\right|^2}{V_0(t,Y^y_t)} \le \alpha_t V_0(t,Y^y_t) + \beta_t \tag{4.2.4}$$

for all $t\in[0,T]$, $y\in O$ and stopping times τ, where L is given by (4.2.2) with $\tilde{b},\tilde{\sigma}$ replacing b,σ.

Definition 4.2.2. A function V is referred to as a Lyapunov function if there exist $\bar{n}\in\mathbb{N}$, $p^*\in[1,\infty)$, open $\bar{O}\subseteq\mathbb{R}^{\bar{n}}$, $\tilde{b}^\cdot_\cdot : \Omega\times[0,T]\times\bar{O}\to\mathbb{R}^{\bar{n}}$, $\tilde{\sigma}^\cdot_\cdot : \Omega\times[0,T]\times\bar{O}\to\mathbb{R}^{\bar{n}\times\bar{n}}$, $V_0\in C^{1,2}([0,T]\times\bar{O})$, along with some $\alpha_\cdot$ and $\beta_\cdot$ such that $V : \Omega\times[0,T]\times\bar{O}\to(0,\infty)$ is a $(\tilde{b}_\cdot, \tilde{\sigma}_\cdot, \alpha_\cdot, \beta_\cdot, p^*, V_0)$-Lyapunov function and

$$\left\| e^{\int_0^T |\alpha_u| du} \right\|_{L^{\frac{p^*}{p^*-1}}(\mathbb{P})} dt + \int_0^T \left\| \frac{\beta_v}{e^{\int_0^v \alpha_u du}} \right\|_{L^{p^*}(\mathbb{P})} dv dt < \infty. \tag{4.2.5}$$

Remark 4.2.1. (i) Smooth functions V satisfying $LV \leq CV$ for some constant C as in [111, Theorem 3.5] form Lyapunov functions with $p^* = 1$, $\alpha_t = C$ and $\beta_t = 0$.

(ii) Lyapunov functions satisfy the stochastic Grönwall inequality as in Theorem 2.4 in [101] along with a finiteness condition on the associated processes, which are properties that will be used many times throughout the chapter.

The following property allows control across families of Lyapunov functions.

Definition 4.2.3. For a family of functions $(\hat{W}_s)_{s\in[0,T]}$, we say that $(\hat{W}_s)_{s\in[0,T]}$ is $(\bar{n}, \bar{O}, V_0)$-local in s if $\bar{n} \in \mathbb{N}$, open $\bar{O} \subseteq \mathbb{R}^{\bar{n}}$, $V_0 \in C^{1,2}([0,\infty) \times \bar{O})$ are such that there exists a constant $C > 0$ satisfying that for any $s \in [0,T]$, $\hat{W}_s : \Omega \times [0,T] \times \bar{O} \to (0,\infty)$ is a $(\tilde{b}^{s,T}_\cdot, \tilde{\sigma}^{s,T}_\cdot, \alpha^{s,T}_\cdot, \beta^{s,T}_\cdot, p^{s,T}, V_0(s+\cdot,\cdot)|_{[0,T]\times\bar{O}})$-Lyapunov function for some $\tilde{b}^{s,T}_\cdot$, $\tilde{\sigma}^{s,T}_\cdot$ together with some $\alpha^{s,T}_\cdot$, $\beta^{s,T}_\cdot$, $p^{s,T}$ where (4.2.5) holds uniformly with bound C, that is,

$$\left\| e^{\int_0^T |\alpha_u^{s,T}| du} \right\|_{L^{\frac{p^{s,T}}{p^{s,T}-1}}(\mathbb{P})} + \int_0^T \left\| \frac{\beta_v^{s,T}}{e^{\int_0^v \alpha_u^{s,T} du}} \right\|_{L^{p^{s,T}}(\mathbb{P})} dv < C. \tag{4.2.6}$$

We say that $(W_s)_{s\in[0,T]}$ is local in s if there exist $\bar{n}, \bar{O}, V_0$ such that $(W_s)_{s\in[0,T]}$ is $(\bar{n}, \bar{O}, V_0)$-local in s.

A family of Lyapunov functions being local in s allows terms of the form $\mathbb{E}[W_s(t, X_t^{s,x})]$ to be bounded uniformly in s after applying Theorem 2.4 in [101]. This is an important property for twice differentiable solutions to Kolmogorov equations, since such solutions and many lemmatic terms depend on a time variable via the starting times s. On the other hand, such a property is in all of the examples mentioned here easily satisfied.

4.3 Moment estimates on derivative processes

The following assumption states our main requirement on the Lyapunov function. Alternative assumptions for the main results in the case where b and σ are independent of t and admit locally Lipschitz derivatives are given in Theorem 4.5.2.

Assumption 7. There exists $G : \Omega \times [0,T] \times O \to [0,\infty)$ such that G is $\mathcal{F} \otimes \mathcal{B}([0,T]) \otimes \mathcal{B}(O)$-measurable and $G(t,\cdot)$ is $\mathcal{F}_t \otimes \mathcal{B}(O)$-measurable,

$$|b(t,x) - b(t,y)| \leq (G(t,x) + G(t,y))|x-y|, \tag{4.3.1}$$

$$\|\sigma(t,x) - \sigma(t,y)\|^2 \leq (G(t,x) + G(t,y))|x-y|^2, \tag{4.3.2}$$

for all $t \in [0,T]$, $x, y \in O$ and such that for any $s \in [0,T]$, there exist locally bounded functions $M : (0,\infty) \to (0,\infty)$, $\bar{x}$ and Lyapunov function V satisfying

$$G(s+t, X_t^{s,x}) \leq m \log V(t, \bar{x}(x)) + M(m) \qquad \mathbb{P}\text{-a.s.} \tag{4.3.3}$$

for all $m > 0$, $x \in O$ and stopping times $t \leq T - s$.

By and large and throughout the chapter, the process Y_t associated with Lyapunov functions can be thought of to be equal to X_t and in the applications here, it is enough to take $G \leq m \log V_0 + M$ in place of (4.3.3); the generality is justified by a trick to increase the set of Lyapunov functions, as exemplified by the inclusion of $\bar{U}$ in [101, Corollary 3.3]. Note that we may just as easily prove with the same effect most of the results in the sequel by introducing a weaker condition that is

$$\int_0^t G(s+r, X_r^{s,x})dr \leq M(m) + m \sum_{i \in I_0} \int_0^t \log V_i(r, \bar{x}_i(x))dr + \log V_i(t, \bar{x}_i(x)) \qquad \mathbb{P}\text{-a.s.}$$

in place of (4.3.3), where I_0 is some finite set. This in some sense covers the inclusion of $\bar{U}$ as mentioned and also the conditions with multiple Lyapunov-type functions as in [44, Theorem 2.24]. In this thesis, we work with (4.3.3) for simplicity.

Assumption 7 is strictly weaker than assuming globally Lipschitz coefficients, since polynomial Lyapunov functions are easily constructed in that case. In addition, whenever continuous differentiability up to some order m^* of b and σ is assumed, we also assume $\mathbb{P}$-a.s.

$$\sum_{\theta \in \mathbb{N}_0^n; |\theta| \leq m^*} \int_0^T \sup_{|x| \leq R} (|\partial^\theta b(t,x)| + \|\partial^\theta \sigma(t,x)\|)dt < \infty, \qquad \forall R > 0. \tag{4.3.4}$$

For $x \in O$, $s \in [0,T]$, let $X_{t(\kappa)}^{s,x}$ be the first t-uniform derivatives in probability of $X_t^{s,x}$ with respect to the initial value in any directions $\kappa \in \mathbb{R}^{2n}$, that is, for any $\epsilon > 0$, $T > 0$, it holds that

$$\mathbb{P}\left(\sup_{t \in [0,T-s]} \left| \frac{X_t^{s,x+r\kappa} - X_t^{s,x}}{r} - X_{t(\kappa)}^{s,x} \right| > \epsilon \right) \to 0$$

as $r \to 0$ with $r \neq 0$, $x + r\kappa \in O$. If $b(t,\cdot)$ and $\sigma(t,\cdot)$ are once continuously differentiable on O for all $t \in [0,\infty)$ and satisfy (4.3.4) with $m^* = 1$, then by Theorem 4.10 in [114], $X_{t(\kappa)}^{s,x}$ exists for any $x \in O$, $s \in [0,T]$ and satisfies the system obtained by formal differentiation of (4.2.1), that is,

$$dX_{t(\kappa)}^{s,x} = (X_{t(\kappa)}^{s,x} \cdot \nabla) b(s+t, X_t^{s,x})dt + (X_{t(\kappa)}^{s,x} \cdot \nabla)\sigma(s+t, X_t^{s,x})dW_t. \tag{4.3.5}$$

By induction, if $b(t,\cdot)$ and $\sigma(t,\cdot)$ are continuously differentiable on O up to some order p for all $t \in [0,\infty)$ and satisfy (4.3.4) with $m^* = p$, then the p^{th}-order t-uniform derivative in probability of $X_t^{s,x}$ with respect to the initial value in directions $(\kappa_i)_{1\le i\le p}$, $\kappa_i \in \mathbb{R}^n$, $|\kappa_i| = 1$, $1 \le i \le p$ exists for any $x \in O$, $s \in [0,T]$ and satisfies the system obtained by a correponding p^{th}-order formal differentiation of (4.2.1).
First we state a straightforward application of the Lyapunov property to obtain an estimate of a time integral, which will be used later and is also demonstrative for many similar derivations in the following. Throughout and consistent with Assumption 7, we omit in the notation the dependence of V, $\bar{x}$ and M on s.

Lemma 4.3.1. *Under Assumption 7, for any constant $c > 0$, it holds that*

$$\begin{aligned}\int_0^{T-s} \mathbb{E}e^{cG(s+t,X_t^{s,x})}dt &\le e^{\hat{M}}\int_0^{T-s}\mathbb{E}[V(t,\bar{x}(x))]dt\\ &\le e^{\hat{M}}\left(\int_0^{T-s}\left\|e^{\int_0^t \alpha_s ds}\right\|_{L^{\frac{p^*}{p^*-1}}(\mathbb{P})}\left(\mathbb{E}[V(0,\bar{x}(x))] + \int_0^t \left\|\frac{\beta_v}{e^{\int_0^v \alpha_u du}}\right\|_{L^{p^*}(\mathbb{P})} dv\right)dt\right)\\ &\le C(\mathbb{E}[V(0,\bar{x}(x))]+1) < \infty\end{aligned}$$

for all $x \in O$ and $T > 0$, where $\hat{M} = M(\frac{1}{c})$ and V is a $(\bar{b}_\cdot^\cdot, \bar{\sigma}_\cdot^\cdot, \alpha_\cdot, \beta_\cdot, p^*, \bar{V}_0)$-Lyapunov function for some $\bar{b}_\cdot^\cdot$, $\bar{\sigma}_\cdot^\cdot$, p^*, $\bar{V}_0$.

Proof. The first inequality is (4.3.3) with $m = \frac{1}{c}$; the second inequality follows by Theorem 2.4 in [101] with $q_1 = 1$, $q_2 = \frac{p^*}{p^*-1}$, $p = p^*$; the third and last inequalities follow by (4.2.5). ☐

Lemma 4.3.2. *Under Assumption 7, for any $k > 0$, $s \in [0,T]$, there exists $\rho > 0$ such that*

$$\mathbb{E}\sup_{0\le t\le T-s}\left|X_{t(\kappa)}^{(r)}\right|^k \le \rho W(x,r\kappa)|r|^k \tag{4.3.6}$$

for all $x \in O$, $r \in \mathbb{R}\backslash\{0\}$, $\kappa \in \mathbb{R}^n$, $|\kappa| = 1$, $x + r\kappa \in O$, where $X_{t(\kappa)}^{(r)} := X_t^{s,x+r\kappa} - X_t^{s,x}$ and $W(x,r\kappa) := 1 + V(0,\bar{x}(x+r\kappa)) + V(0,\bar{x}(x))$. If in addition $b(t,\cdot)$, $\sigma(t,\cdot)$ are continuously differentiable for all $t \ge 0$ and (4.3.4) holds with $m^* = 1$, then

$$\mathbb{E}\sup_{0\le t\le T-s}\left|X_{t(\kappa)}^{s,x}\right|^k \le \rho W(x,0) \tag{4.3.7}$$

$$\lim_{0\ne r\to 0}\mathbb{E}\sup_{0\le t\le T-s}\left|X_{t(\kappa)}^{s,x} - r^{-1}X_{t(\kappa)}^{(r)}\right|^k = 0 \tag{4.3.8}$$

for all $x \in O$, $\kappa \in \mathbb{R}^n$ with $|\kappa| = 1$. If V is local in s, then ρ is independent of s.

Remark 4.3.1. In the proof of Lemma 4.3.2, only a one sided Lipschitz version of (4.3.1) is necessary, see also Section 3.3 in [101] or Corollary 2.31 in [44], which gives similar estimates to (4.3.6).

Proof. For any r,

$$dX^{(r)}_{t(\kappa)} = (b(s+t, X^{s,x+r\kappa}_t) - b(s+t, X^{s,x}_t))dt + (\sigma(s+t, X^{s,x+r\kappa}_t) - \sigma(s+t, X^{s,x}_t))dW_t. \tag{4.3.9}$$

Since X_t is almost surely continuous in t, it holds that $\int_0^t (G(s+u, X^{s,x+r\kappa}_u) + G(s+u, X^{s,x}_u))du \le C\int_0^t (\log V(u, \bar{x}(x+r\kappa)) + \log V(u, \bar{x}(x)) + 1)du < \infty$ for any $\omega \in \Omega$ and finite $0 < t \le T-s$ (for everywhere continuous modifications of X_t), therefore Corollary 2.5 in [101] can be applied with

$$\begin{aligned}
&a_t = b(s+t, X^{s,x+r\kappa}_t) - b(s+t, X^{s,x}_t),\ b_t = \sigma(t, X^{s,x+r\kappa}_t) - \sigma(t, X^{s,x}_t),\\
&\alpha_t = \left(\frac{1}{2} + k \vee 1\right)(G(s+t, X^{s,x+r\kappa}_t) + G(s+t, X^{s,x}_t)),\\
&p = 2k \vee 2,\ \beta_t = 0,\ q_1 = k,\ q_2 = 3k,\ q_3 = \frac{3k}{2},
\end{aligned}$$

to obtain

$$\mathbb{E} \sup_{0\le t\le T-s} \left| X^{(r)}_{t(\kappa)} \right|^k \le C\left(\mathbb{E} e^{\int_0^{T-s} 3k(\frac{1}{2}+k\vee 1)(G(s+u, X^{s,x+r\kappa}_u) + G(s+u, X^{s,x}_u))du} \right)^{\frac{1}{3}} |r|^k. \tag{4.3.10}$$

By Jensen's inequality and Lemma 4.3.1, the expectation on the right-hand side of (4.3.10) satisfies the bound

$$\begin{aligned}
&\mathbb{E} e^{\int_0^{T-s} 3k(\frac{1}{2}+k\vee 1)(G(s+u, X^{s,x+r\kappa}_u) + G(s+u, X^{s,x}_u))du}\\
&\quad\le (T-s)^{-1} e^{2M((6k(\frac{1}{2}+k\vee 1)(T-s))^{-1})} \cdot \mathbb{E} \int_0^{T-s} e^{\frac{1}{2}\log V(u, \bar{x}(x+r\kappa)) + \frac{1}{2}\log V(u, \bar{x}(x)))} du\\
&\quad\le (2(T-s))^{-1} e^{2M((6k(\frac{1}{2}+k\vee 1)(T-s))^{-1})} \mathbb{E} \int_0^{T-s} \Big(V(u, \bar{x}(x+r\kappa)) + V(u, \bar{x}(x)) \Big) du\\
&\quad\le C e^{2M((6k(\frac{1}{2}+k\vee 1)(T-s))^{-1})} (1 + V(0, \bar{x}(x+r\kappa)) + V(0, \bar{x}(x))).
\end{aligned}$$

which easily gives (4.3.6).

The statement for $X_{s(\kappa)}$ follows along the same lines, where instead $X_{s(\kappa)}$ satisfies (4.3.5) and Corollary 2.5 in [101] can be applied as above except with

$$\begin{aligned}
&\alpha_t = (1+2k)(m \log V(t, \bar{x}(x)) + M(m)) > 0 \\
&m = (6k(1+2k)(T-s))^{-1}.
\end{aligned} \tag{4.3.11}$$

Equation (4.3.8) is a known consequence; it is immediate from the definition of $X_{u(\kappa)}$, the previous bounds and

$$\begin{aligned}\mathbb{E}[S^{k_1}] &\leq \epsilon\mathbb{P}(S^{k_1} \leq \epsilon) + \mathbb{E}[\mathbb{1}_{\{S^{k_1}>\epsilon\}}S^{k_1}] \\ &\leq \epsilon\mathbb{P}(S^{k_1} \leq \epsilon) + \mathbb{E}[\mathbb{1}_{\{S^{k_1}>\epsilon\}}]\mathbb{E}[S^k]^{\frac{k_1}{k}} \qquad (4.3.12)\end{aligned}$$

with $S = \sup_{0\leq u\leq T-s}\left|X^{s,x}_{u(\kappa)} - r^{-1}X^{(r)}_{u(\kappa)}\right|$. The final assertion follows by noting that the constants C above are independent of s. $\square$

The following Assumption 8 states our requirements on the higher derivatives of b and σ for results on the higher derivatives of solutions to (4.2.1).

Assumption 8. There exist $p \in \mathbb{N}_0$ such that $b(t,\cdot)|_{\hat{O}}, \sigma(t,\cdot)|_{\hat{O}} \in C^p$ for all $t \geq 0$, $\omega \in \Omega$ and inequality (4.3.4) holds with $m^* = p$. Moreover, for all $s \in [0,T]$ and $k \geq 2$, there exist $M' > 0$, $\hat{n}_k \in \mathbb{N}$, open $\hat{O}_k \subset \mathbb{R}^{\bar{n}_k}$, mappings $\hat{x}_k : O \to \hat{O}_k$ and Lyapunov function $\hat{V}^{s,T}_k : \Omega\times[0,T-s]\times\hat{O}_k \to (0,\infty)$ satisfying for any $x, x' \in O$ and multiindices α with $2 \leq |\alpha| \leq p$ that $\mathbb{P}$-a.s.

$$\begin{aligned}&\left|\partial^\alpha b(s+t, \lambda X^{s,x}_t + (1-\lambda)X^{s,x'}_t)\right| + \|\partial^\alpha\sigma(s+t,\lambda X^{s,x}_t + (1-\lambda)X^{s,x'}_t)\|^2 \\ &\quad\leq M'(1 + \hat{V}^{s,T}_k(t,\hat{x}_k(x)) + \hat{V}^{s,T}_k(t,\hat{x}_k(x')))^{\frac{1}{k}} \qquad (4.3.13)\end{aligned}$$

holds for all $t \in [0, T-s]$, $\lambda \in [0,1]$.

In the following, for $\kappa = (\kappa_i)_{1\leq i\leq l}$, $\kappa_i \in \mathbb{R}^n$, the l^{th} order t-uniform derivatives in probability of a process Z^x_t with respect to initial condition x in the directions $\kappa_1,\ldots,\kappa_l \in \mathbb{R}^n$ is denoted by $\partial^{(\kappa)}Z^x_t$.

Theorem 4.3.3. *Under Assumptions 7 and 8, for any $s \in [0,T]$, constants $1 \leq l \leq p-1$, $k_1 > 0$, there exist $i^* \in \mathbb{N}$, $\nu \geq \frac{k_1}{2}$, $\{l_i\}_{i\in\{1,\ldots,i^*\}} \subset (0,\infty)$ and a finite order polynomial q_0, the degree of which is independent of $s, T, V, \hat{V}^{s,T}_k$, such that*

$$\mathbb{E}\sup_{0\leq t\leq T-s}\left|\partial^{(\kappa)}X^{s,x+r\kappa_{l+1}}_t - \partial^{(\kappa)}X^{s,x}_t\right|^{k_1} \leq (T-s)^\nu q(x, x+r\kappa_{l+1})|r|^{k_1} \qquad (4.3.14)$$

$$\mathbb{E}\sup_{0\leq t\leq T-s}\left|\partial^{(\bar{\kappa})}X^{s,x}_t\right|^{k_1} \leq (T-s)^\nu q(x,x) \qquad (4.3.15)$$

$$\lim_{r\to 0}\mathbb{E}\sup_{0\leq t\leq T-s}\left|\partial^{(\bar{\kappa})}X^{s,x}_t - r^{-1}(\partial^{(\kappa)}X^{s,x+r\kappa_{l+1}}_t - \partial^{(\kappa)}X^{s,x}_t)\right|^{k_1} = 0 \qquad (4.3.16)$$

for all initial condition $x \in O$, $r \in \mathbb{R}\backslash\{0\}$, $\kappa_i \in \mathbb{R}^n$, $|\kappa_i| = 1$, $1 \leq i \leq l+1$, $x + r\kappa_{l+1} \in O$, *where $\kappa = (\kappa_i)_{1\le i\le l}$, $\bar\kappa = (\kappa_i)_{1\le i\le l+1}$ and $q : O \times O \to \mathbb{R}$ is given by*

$$q(y,y') = q_0(V(0,\bar x(y)), (\hat V^{s,T}_{l_i}(0,\hat x_{l_i}(y)))_{i\in\{1,\dots,i^*\}}, (\hat V^{s,T}_{l_i}(0,\hat x_{l_i}(y')))_{i\in\{1,\dots,i^*\}}). \quad (4.3.17)$$

If V and $\hat V^{s,T}_k$ are local in s for every k, then the form of the polynomial q_0 is independent of $s, T, V, \hat V^{s,T}_k$.

Remark 4.3.2. Assumption 8 can be weakened if only finite order moments of the derivatives in Theorem 4.3.3 are sought after, that is, if the statements in Theorem 4.3.3 are only required to hold for k_1 up to some finite $k_1 \le K_1$. In particular, $M', \hat n_k, \hat O_k, \hat x_k, \hat V_k$ in Assumption 8 need only exist for k up to some finite $k \le K$. In the same vein, Assumption 7 can be weakened (so that the Lipschitz constants are only required to be $O(\log V)$ and $O(\sqrt{\log V})$) in this case if T is sufficiently small.

Proof. Fix $k_1 > 0$, $s \in [0,T]$, let J be the set of strictly increasing functions from $\mathbb{N}$ to itself and $D^{(\kappa)} b(s+t, X^{s,x}_t)$ denote the formal derivative of $b(s+t, X^{s,x}_t)$ with respect to x in the directions indicated by κ. In particular,

$$\begin{aligned} D^{(\kappa)} b(s+t, X^{s,x}_t) &= \Big(\partial^{(\kappa)} X^x_t \cdot \nabla\Big) b(s+t, X^{s,x}_t) \\ &\quad + q_{b,X^{s,x}_t}\Bigg(\bigg(\prod_{1\le i\le l'} \partial^{(\kappa_{j(i)})}\bigg) X^{s,x}_t, 1 \le l' \le l-1, j \in J\Bigg), \end{aligned}$$

where the last term denotes a $\mathbb{R}^n$-valued polynomial taking arguments as indicated, for which exactly l of the operators $\partial^{(\kappa_i)}$ appear in each term and coefficients are spatial derivatives between orders 2 and l of elements of b evaluated at $(s+t, X^{s,x}_t)$. The term $D^{(\kappa)}\sigma(s+t, X^{s,x}_t)$ is similarly defined. Denoting $x' = x + r\kappa_{l+1}$, the difference processes of the derivatives satisfy

$$\begin{aligned} d(\partial^{(\kappa)} X^{s,x'}_t - \partial^{(\kappa)} X^{s,x}_t) &= (D^{(\kappa)} b(s+t, X^{s,x'}_t) - D^{(\kappa)} b(s+t, X^{s,x}_t))dt \\ &\quad + (D^{(\kappa)}\sigma(s+t, X^{s,x'}_t) - D^{(\kappa)}\sigma(s+t, X^{s,x}_t))dW_t \end{aligned}$$

on $t \in [0, T-s]$ for all $x, x' \in O$, $r \in \mathbb{R}\setminus\{0\}$, $\kappa_i \in \mathbb{R}^n$, $|\kappa_i| = 1$, $1 \le i \le l+1$.
We proceed by strong induction in l for (4.3.14). A base case has been established in Lemma 4.3.2. By the fundamental theorem of calculus on derivatives of b and σ, inequalities (4.3.13), (4.3.1), (4.3.2) and (4.3.3) with $m_1 = ((4k_1 \vee 4) - 1)m_2 = \frac{1}{4T}(\frac{1}{k_1} -$

$\frac{1}{2k_1\vee 2}$), so that $M(m_1), M(m_2)$ are locally bounded as functions of T, it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&\left|D^{(\kappa)}b(s+t,X_t^{s,x'})-D^{(\kappa)}b(s+t,X_t^{s,x})\right|\\
&\quad\le \sum_i\left|(\partial^{(\kappa)}X_t^{s,x'}-\partial^{(\kappa)}X_t^{s,x})_i\right||\partial_i b(s+t,X_t^{s,x})|+H(t,X_t^{s,x},X_t^{s,x'})\hat{q}_t\\
&\quad\le 2\left|\partial^{(\kappa)}X_t^{s,x'}-\partial^{(\kappa)}X_t^{s,x}\right|(m_1\log V(t,\bar{x}(x))+M(m_1))+H(t,X_t^{s,x},X_t^{s,x'})\hat{q}_t,\\
&\|D^{(\kappa)}\sigma(s+t,X_t^{s,x'})-D^{(\kappa)}\sigma(s+t,X_t^{s,x})\|^2\\
&\quad\le 2\sum_i\left|(\partial^{(\kappa)}X_t^{s,x'}-\partial^{(\kappa)}X_t^{s,x})_i\right|^2\|\partial_i\sigma(s+t,X_t^{s,x})\|^2+(H(t,X_t^{s,x},X_t^{s,x'})\hat{q}_t)^2\\
&\quad\le 4\left|\partial^{(\kappa)}X_t^{s,x'}-\partial^{(\kappa)}X_t^{s,x}\right|^2(m_2\log V(t,\bar{x}(x))+M(m_2))+(H(t,X_t^{s,x},X_t^{s,x'})\hat{q}_t)^2,
\end{aligned}$$

on $t\in[0,T]$, where

$$H(t,X_t^{s,x},X_t^{s,x'})=M'\Big(1+\hat{V}_{4k_1\vee 4}(t,\hat{x}_{4k_1\vee 4}(x))+\hat{V}_{4k_1\vee 4}(t,\hat{x}_{4k_1\vee 4}(x'))\Big)^{\frac{1}{4k_1\vee 4}} \quad (4.3.18)$$

and $\hat{q}_s$ denotes a polynomial with constant coefficients taking arguments from the set $S=S_1\cup S_2$,

$$\begin{aligned}
S_1&=\left\{\left|\Bigg(\prod_{1\le i\le l'}\partial^{(\kappa_{j(i)})}\Bigg)X_t\right|:1\le l'\le l, j\in J, X_t\in\{X_t^{s,x'},X_t^{s,x}\}\right\}\\
S_2&=\left\{\left|\Bigg(\prod_{1\le i\le l'}\partial^{(\kappa_{j(i)})}\Bigg)(X_t^{s,x'}-X_t^{s,x})\right|:1\le l'\le l-1, j\in J\right\}\cup\{|X_t^{s,x'}-X_t^{s,x}|\},
\end{aligned}$$

for which exactly l of the operators $\partial^{(\kappa_i)}$ appear in each term of $\hat{q}_s$ and a factor from S_2 appears exactly once in each term. Note for $p\ge 2$ and by Lemma 4.3.1, it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&\int_0^{T-s}|(m_1+(p-1)m_2)\log V(t,\bar{x}(x))+M(m_1)+(p-1)M(m_2)|dt\\
&\quad<\int_0^{T-s}|(m_1+(p-1)m_2)V(t,\bar{x}(x))+M(m_1)+(p-1)M(m_2)|dt<\infty
\end{aligned}$$

on $t \in [0, T-s]$. Corollary 2.5 in [101] can then be applied with

$$\begin{aligned}
a_t &= D^{(\kappa)} b(s+t, X_t^{s,x'}) - D^{(\kappa)} b(s+t, X_t^{s,x}),\\
b_t &= D^{(\kappa)} \sigma(s+t, X_t^{s,x'}) - D^{(\kappa)} \sigma(s+t, X_t^{s,x}),\\
\alpha_t &= 2(m_1 + (p-1)m_2) \log V(t, \bar{x}(x)) + 2(M(m_1) + (p-1)M(m_2)) + \frac{1}{2} > 0, \quad (4.3.19)\\
\beta_t &= \sqrt{2k_1 \vee 2} H(t, X_t^{s,x}, X_t^{s,x'}) \hat{q}_t,\\
p &= 4k_1 \vee 4,\ q_1 = k_1,\ q_2 = \left(\frac{1}{k_1} - \frac{1}{2k_1 \vee 2}\right)^{-1},\ q_3 = 2k_1 \vee 2
\end{aligned}$$

to obtain

$$\mathbb{E} \sup_{0 \le t \le T-s} \left| \partial^{(\kappa)} X_t^{s,x'} - \partial^{(\kappa)} X_t^{s,x} \right|^{k_1} \le C A_{T-s}^{(1)} A_{T-s}^{(2)},$$

where

$$\begin{aligned}
A_{T-s}^{(1)} &:= \left(\mathbb{E}\left[e^{\int_0^{T-s} (2q_2(m_1+(p-1)m_2) \log V(u, \bar{x}(x)) + 2q_2(M(m_1)+(p-1)M(m_2)) + \frac{q_2}{2}) du} \right] \right)^{\frac{k_1}{q_2}}\\
A_{T-s}^{(2)} &:= \left(\mathbb{E}\left[\int_0^{T-s} (2k_1 \vee 2)(H(u, X_u^{s,x}, X_u^{s,x'}) \hat{q}_u)^2 du \right]^{\frac{q_3}{2}} \right)^{\frac{k_1}{q_3}}.
\end{aligned}$$

By substituting our expressions for q_2, m_1, m_2, setting $V(u, \bar{x}(x)) = 1$ for all $u \ge T-s$ and using Lemma 4.3.1, the first expectation has the bound

$$\begin{aligned}
&A_{T-s}^{(1)}\\
&\le \left(\mathbb{E}\left[\frac{1}{T-s} \int_0^{T-s} e^{(T-s)(2q_2(m_1+(p-1)m_2) \log V(u, \bar{x}(x)) + 2q_2(M(m_1)+(p-1)M(m_2)) + \frac{q_2}{2})} du \right] \right)^{\frac{k_1}{q_2}}\\
&\le C \left(\mathbb{E}\left[\frac{1}{T-s} \int_0^{T-s} V(u, \bar{x}(x)) du \right] \right)^{\frac{k_1}{q_2}}\\
&\le C(\mathbb{E}[V(0, \bar{x}(x))] + 1)^{\frac{k_1}{q_2}},
\end{aligned}$$

where note C is, here and in the rest of the proof, locally bounded as a function of T and also of s, T if V is local in s. On the other hand, by the inductive argument and the form of H, $\hat{q}_s$ and q_3, it holds that

$$\begin{aligned}
A^{(2)}_{T-s} &\leq C\bigg(\mathbb{E}\bigg[\bigg(\int_0^{T-s} H(u,X_u^{s,x},X_u^{s,x'})^2du\bigg)^{k_1\vee 1}\sup_{0\leq u\leq T-s}\hat{q}_u^{2k_1\vee 2}\bigg]\bigg)^{\frac{k_1}{2k_1\vee 2}}\\
&\leq C\bigg(\mathbb{E}\bigg[\int_0^{T-s} H(u,X_u^{s,x},X_u^{s,x'})^2du\bigg]^{2k_1\vee 2}\bigg)^{\frac{k_1}{4k_1\vee 4}}\bigg(\mathbb{E}\sup_{0\leq u\leq T-s}\hat{q}_u^{4k_1\vee 4}\bigg)^{\frac{k_1}{4k_1\vee 4}}\\
&\leq C\bigg((T-s)^{(2k_1\vee 2)-1}\int_0^{T-s}\mathbb{E}(1+\hat{V}_{4k_1\vee 4}(u,\hat{x}_{4k_1\vee 4}(x))\\
&\qquad +\hat{V}_{4k_1\vee 4}(u,\hat{x}_{4k_1\vee 4}(x')))du\bigg)^{\frac{k_1}{4k_1\vee 4}}\tilde{q}(x,x')|r|^{k_1},
\end{aligned}$$

where $\tilde{q}(x,x')=\tilde{q}_0(V(0,\bar{x}(x)),(\hat{V}_{l_i}(0,\hat{x}_{l_i}(x)))_{i\in\{1,\dots,\hat{i}^*\}},(\hat{V}_{l_i}(0,\hat{x}_{l_i}(x')))_{i\in\{1,\dots,\hat{i}^*\}})$ for some $\hat{i}^*\in\mathbb{N}$, $\{l_i\}_{i\in\{1,\dots,\hat{i}*\}}\subset(0,\infty)$ and finite order polynomial $\tilde{q}_0$ taking arguments as indicated. Therefore, by Theorem 2.4 in [101] with $q_1=1$, it holds that

$$\begin{aligned}
A^{(2)}_{T-s} &\leq C\bigg((T-s)^{(2k_1\vee 2)}(1+\hat{V}_{4k_1\vee 4}(0,\hat{x}_{4k_1\vee 4}(x))+\hat{V}_{4k_1\vee 4}(0,\hat{x}_{4k_1\vee 4}(x')))\bigg)^{\frac{k_1}{4k_1\vee 4}}\\
&\qquad\cdot\tilde{q}(x,x')|r|^{k_1},
\end{aligned}$$

which concludes the proof for (4.3.14). Inequality (4.3.15) follows along the same lines, therefore the argument is not repeated. Equation (4.3.16) holds by (4.3.12) with

$$S=\sup_{0\leq u\leq t}\Big|\partial^{(\bar{\kappa})}X_u^{s,x}-r^{-1}(\partial^{(\kappa)}X_u^{s,x'}-\partial^{(\kappa)}X_u^{s,x})\Big|.$$

□

Remark 4.3.3. In the proofs of Lemma 4.3.2 and Theorem 4.3.3, the stochastic Grönwall inequality, that is, Theorem 2.4 and Corollary 2.5 in [101], can be replaced with Lemma 4.2 in [114] and Theorem 3.5 in [111]. For this, one works directly with the SDEs governing $\big|\partial^{(\kappa)}X_t\big|^{k_1}$ in the proof and inequality (4.2.4) is to be replaced by $(\partial_t+L)V_0\leq CV_0$. The latter point raises complications in the examples of [44] with domain not equal to $\mathbb{R}^n$.

4.4 Kolmogorov's equation

Throughout this section, we assume that b and σ are nonrandom functions. In Section 4.4.1, the moment estimates from Section 4.3 are used to derive p^{th} differentiability of a Feynman-Kac semigroup and in particular of $x\mapsto\mathbb{E}g(X_t^{s,x})$ for $X_t^{s,x}$ solving (4.2.1). We allow the functions such as g to be bounded by Lyapunov functions. As such, the proof, although the approach of which is classical, demands finer attention compared to [114] for example, in which g and its derivatives are only required to be polynomially bounded. This regularity is then used to show that the semigroup solves the Kolmogorov equation in the almost everywhere sense in Section 4.4.2. In Section 4.4.3, we complement our results with a criterion for the Kolmogorov equation to be solved in the distributional sense. The criterion is based on the maximal dissipativity of the closure of the infinitesimal generator associated to (4.1.1).

4.4.1 Semigroup differentiability

To begin, we state a condition that will be imposed on functions such as g.

Definition 4.4.1. For $p \in \mathbb{N}$, $k > 1$, $h : \Omega \times [0,T] \times \hat{O} \to \mathbb{R}$ with $h(t,\cdot) \in C^p(\hat{O})$ for all $(\omega,t) \in \Omega \times [0,T]$, we say that h has Lyapunov derivatives up to order (p,k) if there exist $(V^{s,T})_{s\in[0,T]}$ local in s, locally bounded $\tilde{x}$ and constant $N > 0$ such that for any $s \in [0,T]$ and multiindices α with $0 \leq |\alpha| \leq p$, it holds $\mathbb{P}$-a.s. that

$$\left|\partial^\alpha h(s+t, \lambda X_t^{s,x} + (1-\lambda) X_t^{s,x'})\right| \leq N(1 + V^{s,T}(t,\tilde{x}(x)) + V^{s,T}(t,\tilde{x}(x')))^{\frac{1}{k}} \tag{4.4.1}$$

for all stopping times $t \leq T - s$, $x, x' \in O$ and $\lambda \in [0,1]$.

We make the following mild assumptions about the Lyapunov functions $V^{s,T}$ associated to Lyapunov derivatives in Definition 4.4.1 or otherwise. These are gathered with additional assumptions on the SDE (4.2.1). Assumption 9 will be referenced only when $V^{s,T}$ has been given in the context.

Assumption 9. For each $R \geq 0$, there exists a Borel, locally integrable $K_\cdot(R) : [0,\infty) \to [0,\infty)$ such that

$$2\langle x - y, b(t,x) - b(t,y)\rangle + \|\sigma(t,x) - \sigma(t,y)\|^2 \leq K_t(R)|x-y|^2$$

for all $t \geq 0$, $x, y \in B_R \cap O$. For any $s \geq 0$, $T > 0$, $x \in O$, there exists a $\mathbb{P}$-a.s. continuous O-valued unique solution $X_t^{s,x}$ to (4.2.1) on $[0,T]$. Moreover, for any $T > 0$, there exist $\bar{n} \in \mathbb{N}$, open $\bar{O} \subseteq \mathbb{R}^{\bar{n}}$, $V_0 \in C^{1,2}([0,\infty)\times\bar{O})$, $\tilde{x} : \mathbb{R}^n \to \bar{O}$, $\hat{G} : [0,\infty)\times\mathbb{R}^n \to \mathbb{R}$, constant $C \geq 0$ and $0 < \bar{l} \leq 1$ such that

(i) $(V^{s,T})_{s\in[0,T]}$ is $(\bar{n}, \bar{O}, V_0)$-local in s,

(ii) for any $s \geq 0$, $V^{s,T}$ is a $(\tilde{b}_\cdot, \tilde{\sigma}_\cdot, \alpha_\cdot, C, p^*, \tilde{V}_0)$-function for some $\tilde{b}_\cdot, \tilde{\sigma}_\cdot, \alpha_\cdot, p^*, \tilde{V}_0$,

(iii) for any $s \geq 0$, it holds $\mathbb{P}$-a.s. that

$$V^{s+\tau,T}(0, \tilde{x}(X_\tau^{s,x}))^{\bar{l}} \leq C(1 + V^{s,T}(\tau, \tilde{x}(x)))$$

for all $x \in O$ and stopping times $\tau \leq T$,

(iv) for any $s \geq 0$, it holds that $\lim_{|x|\to\infty} \inf_{t\in[0,T]} \hat{G}(t,x) = \infty$ and $\mathbb{P}$-a.s. that

$$\hat{G}(s+t, X_t^{s,x}) \leq V^{s,T}(t, \tilde{x}(x))$$

for all $t \in [0,T]$, $x \in O$.

Beside the first two sentences, Assumption 9 is satisfied by the Lyapunov functions considered for example in [44, Corollary 2.4]. More specifically, taking α and the functions U, $\bar{U}$ from there, for $\bar{n} = n+1$, one may take $V_0 = V_0(t,(x,y)) = e^{U(x)e^{-\alpha t}+y}$ and $\tilde{x} = \tilde{x}(x) = (x,0) \in \mathbb{R}^{n+1}$, then

$$\tilde{b}(t,(x,y)) = (b(t,x), \bar{U}(t,x)), \quad \tilde{\sigma}(t,(x,y)) = \begin{pmatrix} \sigma(t,x) & 0 \\ 0 & 0 \end{pmatrix}, \quad \hat{G}(t,x) = e^{U(x)e^{-\alpha t}}$$

for $t \geq 0$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}$ and the latter statements of Assumption 9 are satisfied by the conditions on U and $\bar{U}$ if $\lim_{|x|\to\infty} U(x) = \infty$ and $\bar{U} \geq C$ for some $C \in \mathbb{R}$ everywhere.

Theorem 4.4.2. *Let $T > 0$, let Assumptions 7, 8 hold and let $f : \Omega \times [0,T] \times \mathbb{R}^n \to \mathbb{R}$, $c : \Omega \times [0,T] \times \mathbb{R}^n \to [0,\infty)$, $g : \Omega \times \mathbb{R}^n \to \mathbb{R}$ be such that $f(\cdot,x), c(\cdot,x)$ are $\mathcal{F} \otimes \mathcal{B}([0,T])$-measurable functions for every $x \in \hat{O}$, satisfying $\int_0^T \sup_{x\in B_R\cap\hat{O}}(|c(t,x)|+|f(t,x)|)dt < \infty$ for every $R > 0$ and $\mathbb{P}$-a.s. $f(t,\cdot)|_{\hat{O}}, c(t,\cdot)|_{\hat{O}}, g|_{\hat{O}} \in C^p(\hat{O})$ for all $(\omega,t) \in \Omega \times [0,T]$. Suppose there exists $k_2 > 1$ such that f and g have Lyapunov derivatives up to order (p,k_2). There exists $K > 1$ such that if for any $1 < k' < K$, c has Lyapunov derivatives up to order (p,k') and all of the Lyapunov functions associated to Lyapunov derivatives are such that Assumption 9 is satisfied with $\bar{l} > k_2^{-1}, K^{-1}$, then the following statements hold.*

(i) For u given by

$$u(s,t,x) = \int_0^t f(s+r, X_r^{s,x}) e^{-\int_0^r c(s+w, X_w^{s,x})dw} dr + g(X_t^{s,x}) e^{-\int_0^t c(s+w, X_w^{s,x})dw} \tag{4.4.2}$$

defined for $(s,x) \in [0,T] \times O$ and stopping times $t \leq T-s$, the expectation $\mathbb{E}u(s,t,x)$ is continuously differentiable in x up to order p.

(ii) For every multiindex β with $0 \leq |\beta| \leq p$, there exists a finite order polynomial q^*, the form of which is independent of $\hat{V}^{s,T}$ and $\hat{V}_k^{s,T}$, such that for all stopping times $t \leq T-s$, it holds that

$$|\partial_x^\beta \mathbb{E}u(s,t,x)| \leq q^*(V(0,\bar{x}(x)), V^{s,T}(0,\tilde{x}(x)), \hat{V}_{l_i}^{s,T}(0,\hat{x}_{l_i}(x)) : i \in I^*) \tag{4.4.3}$$

on $(s,x) \in [0,T] \times O$, where $I^* \subset \mathbb{N}$ is finite, $l_i > 0$ and $\hat{x}$, $\hat{x}_{l_i}$, $V^{s,T}$, $\hat{V}^{s,T}_{l_i}$ represent any and all of the functions across $h \in \{f,c,g\}$, multiindices α with $0 \leq |\alpha| \leq |\beta|$ and $k \in K_0 \subset (0,K)$ for some finite K_0.

(iii) If for each k and multiindex α with $0 \leq |\alpha| \leq p$, the mappings $\hat{x}_k$ are independent of s and $(\hat{V}^{s,T}_k)_{s\in[0,T]}$ is local in s, then $|\partial^\alpha_x u|$ is locally bounded for every multiindex α with $0 \leq |\alpha| \leq p$ and if $p \geq 2$, then for any $R > 0$, there exists a constant $N > 0$ such that

$$|\mathbb{E}u(s',T-s',x) - \mathbb{E}u(s,T-s,x)| \leq N|t-s|$$

for all $s,s' \in (0,T)$ and $x \in B_R$.

We prove first a lemma that will used in the proof of Theorem 4.4.2. Throughout the proofs of Theorem 4.4.2, Lemma 4.4.3 and consistent with the statement of the results, we omit in the notation any dependence of $V^{s,T}$, $\tilde{x}$ and k_2 on k and h.

Lemma 4.4.3. *Let the first sentence of Theorem 4.4.2 hold and let c have Lyapunov derivatives up to order (p,k') for all $1 \leq k' < K$, for K from the same theorem. For any $h \in \{f,c,g\}$, $k_3 > 0$ with $k_3 < k_2$ if $h \in \{f,g\}$, $s \in [0,T]$, $x \in O$, $\kappa \in \mathbb{R}^n$ with $|\kappa| = 1$, $\lambda' \in [0,1]$, multiindex α with $0 \leq |\alpha| \leq p$ and stopping time $t \leq T-s$, it holds that*

$$\mathbb{E}\int_0^t \Big|\partial^\alpha h(s+v,\lambda' X^{s,x'}_v + (1-\lambda')X^{s,x}_v) - \partial^\alpha h(s+v,X^{s,x}_v)\Big|^{k_3} dv \to 0 \tag{4.4.4}$$

$$\mathbb{E}\Big|\int_0^1 \partial^\alpha h(s+t,\lambda X^{s,x'}_t + (1-\lambda)X^{s,x}_t)d\lambda - \partial^\alpha h(s+t,X^{s,x}_t)\Big|^{k_3} \to 0$$

$$\mathbb{E}\int_0^t \Big|\int_0^1 \partial^\alpha h(s+v,\lambda X^{s,x'}_v + (1-\lambda)X^{s,x}_t)d\lambda - \partial^\alpha h(s+v,X^{s,x}_v)\Big|^{k_3} dv \to 0$$

as $x' \to x$, where the derivatives ∂^α are in the spatial argument and $g(t,\cdot) = g$.

Proof. For any $\epsilon > 0$, $s \in [0,T]$ and stopping time $t \leq T-s$, note that

$$\mathbb{P}(\sup_{0\leq u\leq T-s}|X^{s,x'}_u - X^{s,x}_u| \leq \epsilon) \leq \mathbb{P}(|X^{s,x'}_t - X^{s,x}_t| \leq \epsilon),$$

so that for any $\lambda \in [0,1]$, by Theorem 1.7 in [114], it holds that $\lambda X^{s,x'}_t + (1-\lambda)X^{s,x}_t - X^{s,x}_t = \lambda(X^{s,x'}_t - X^{s,x}_t) \to 0$ in probability as $x' \to x$ (sequentially). Therefore for any multiindex α, $\hat{J} := \partial^\alpha h(s+t,\lambda X^{s,x'}_t + (1-\lambda)X^{s,x}_t) - \partial^\alpha h(s+t,X^{s,x}_t) \to 0$ in probability by Theorem 20.5 in [17]. Moreover, if $h \in \{f,g\}$, by (4.3.12) with $k_1 = k_3$, $k = k_2$ and $S = |\hat{J}|$, it holds that $\mathbb{E}|\hat{J}|^{k_3} \to 0$ as $x' \to x$. The same holds for $h = c$ using (4.3.12) with some $k > k_3$ instead. By the assumption (4.4.1) and Theorem 2.4 in [101], it holds that

$$\mathbb{E}\Big|\partial^\alpha h(s+u, \lambda X_u^{s,x'} + (1-\lambda)X_u^{s,x})\Big|^{k_3} \leq C\mathbb{E}(1 + V^{s,T}(0, \tilde{x}(x)) + V^{s,T}(0, \tilde{x}(x'))),$$

where C here is independent of r, λ and u, so that Jensen's inequality, Fubini's theorem and dominated convergence theorem concludes the proof. $\square$

Proof of Theorem 4.4.2. For $x \in O$, $s \in [0,T]$, stopping time $t \leq T - s$, $\kappa \in \mathbb{R}^n$, $r \in \mathbb{R} \setminus \{0\}$, $|\kappa| = 1$, let $x' := x + r\kappa \in O$ and for $h \in \{f, c, g\}$, let

$$\begin{aligned}
h'_t &:= \int_0^1 \nabla h(s+t, \lambda X_t^{s,x'} + (1-\lambda)X_t^{s,x})d\lambda, \qquad \hat{h}(t,x) := h(s+t, X_t^{s,x}),\\
c_t^\dagger &:= \int_0^1 e^{-\lambda\int_0^t c(s+u, X_u^{s,x'})du - (1-\lambda)\int_0^t c(s+u, X_u^{s,x})du} d\lambda, \qquad \hat{c}(t,x) := e^{-\int_0^t c(s+u, X_u^{s,x})du},
\end{aligned}$$

where ∇ denotes the gradient in the spatial argument, $g(s+t,\cdot) = g$ and the same for its derivatives. For (i), we show first once directional differentiability. Let $h \in \{f, g\}$; it holds that

$$\begin{aligned}
&\bigg|\frac{\mathbb{E}\hat{h}(t,x')\hat{c}(t,x') - \mathbb{E}\hat{h}(t,x)\hat{c}(t,x)}{r} - \mathbb{E}\bigg[\nabla h(s+t, X_t^{s,x}) \cdot X_{t(\kappa)}^{s,x}\hat{c}(t,x)\\
&\quad - \hat{h}(t,x)\hat{c}(t,x)\int_0^t \nabla c(s+u, X_u^{s,x}) \cdot X_{u(\kappa)}^{s,x} du\bigg]\bigg|\\
&\quad\leq \bigg|\frac{\mathbb{E}\hat{h}(t,x')\hat{c}(t,x') - \mathbb{E}\hat{h}(t,x)\hat{c}(t,x')}{r} - \mathbb{E}h'_t \cdot X_{t(\kappa)}^{s,x}\hat{c}(t,x)\bigg|\\
&\qquad + \bigg|\frac{\mathbb{E}\hat{h}(t,x)\hat{c}(t,x') - \mathbb{E}\hat{h}(t,x)\hat{c}(t,x)}{r} + \mathbb{E}\hat{h}(t,x)c_t^\dagger r^{-1}\bigg(\int_0^t (c(s+u, X_u^{s,x'})\\
&\qquad - c(s+u, X_u^{s,x}))du\bigg)\bigg| + \Big|\mathbb{E}h'_t \cdot X_{t(\kappa)}^{s,x}\hat{c}(t,x) - \mathbb{E}\nabla h(s+t, X_t^{s,x}) \cdot X_{t(\kappa)}^{s,x}\hat{c}(t,x)\Big|\\
&\qquad + \bigg|\mathbb{E}\hat{h}(t,x)c_t^\dagger r^{-1}\bigg(\int_0^t (c(s+u, X_u^{s,x'}) - c(s+u, X_u^{s,x}))du\bigg)\\
&\qquad - \mathbb{E}\hat{h}(t,x)\hat{c}(t,x)\int_0^t \nabla c(s+u, X_u^{s,x}) \cdot X_{u(\kappa)}^{s,x} du\bigg|.
\end{aligned} \tag{4.4.5}$$

The first three terms on the right-hand side of (4.4.5) converge to 0 as $r \to 0$ by the fundamental theorem of calculus, (4.4.1), Lemma 4.3.2 and Lemma 4.4.3. For the last term, Hölder's inequality yields

$$
\begin{aligned}
&\bigg|\mathbb{E}\hat{h}(t,x)c_t^\dagger r^{-1}\bigg(\int_0^t (c(s+u,X_u^{s,x'})-c(s+u,X_u^{s,x}))du\bigg) \\
&\quad - \mathbb{E}\hat{h}(t,x)\hat{c}(t,x)\int_0^t \nabla c(s+u,X_u^{s,x})\cdot X_{u(\kappa)}^{s,x}du\bigg| \\
&\qquad \le \Big\|\hat{h}(t,x)\Big\|_{L^{k_2}(\mathbb{P})}\bigg\|c_t^\dagger\bigg(\int_0^t\bigg(\frac{c(s+u,X_u^{s,x'})-c(s+u,X_u^{s,x})}{r}-\nabla c(s+u,X_u^{s,x}) \\
&\qquad\quad \cdot X_{u(\kappa)}^{s,x}\bigg)du\bigg)+(c_t^\dagger-\hat{c}(t,x))\int_0^t \nabla c(s+u,X_u^{s,x})\cdot X_{u(\kappa)}^{s,x}du\bigg\|_{L^{k_2'}(\mathbb{P})},
\end{aligned}
\tag{4.4.6}
$$

where $\frac{1}{k_2}+\frac{1}{k_2'}=1$. By (4.4.1) and Theorem 2.4 in [101], we have $\mathbb{E}|\hat{h}(t,x)|^{k_2}\le C(1+V^{s,T}(0,\tilde{x}(x))+V^{s,T}(0,\tilde{x}(x')))$. Moreover, Hölder's inequality yields

$$
\begin{aligned}
&\mathbb{E}\bigg|(c_t^\dagger-\hat{c}(t,x))\int_0^t \nabla c(s+u,X_u^{s,x})\cdot X_{u(\kappa)}^{s,x}du\bigg|^{k_2'} \\
&\quad \le \bigg(\mathbb{E}\Big|c_t^\dagger-\hat{c}(t,x)\Big|^{2k_2'}\bigg)^{\frac{1}{2}}\bigg(\mathbb{E}\bigg|\int_0^t \nabla c(s+u,X_u^{s,x})\cdot X_{u(\kappa)}^{s,x}du\bigg|^{2k_2'}\bigg)^{\frac{1}{2}}.
\end{aligned}
\tag{4.4.7}
$$

For the first factor on the right-hand side, note that by (4.4.4) in Lemma 4.4.3, we have $\int_0^t c(s+u,X_u^{s,x'})du\to\int_0^t c(s+u,X_u^{s,x})du$ in probability as $r\to 0$, so that

$$
\hat{S}_t := e^{-\lambda\int_0^t (c(s+u,X_u^{s,x'})-(1-\lambda)\int_0^t c(s+u,X_u^{s,x}))du}-e^{-\int_0^t c(s+u,X_u^{s,x})du}\to 0
$$

in probability by the continuous mapping theorem and $\mathbb{E}|c_t^\dagger-\hat{c}(t,x)|^{2k_2'}\le\int_0^1\mathbb{E}|\hat{S}_t|^{2k_2'}d\lambda\to 0$ as $r\to 0$ by (4.3.12) with $k_1=2k_2'$, $k>2k_2'$ and $S=\hat{S}_t$. By setting $K>2k_2'$, the second factor on the right-hand side of (4.4.7) is clearly bounded independently of r (and of t) by Hölder's inequality, our assumption on the derivatives of c and Lemma 4.3.2. For the remaining term in the second factor on the right-hand side of (4.4.6), the triangle inequality on $L^{k_2'}(\mathbb{P})$ yields

$$
\begin{aligned}
&\bigg\|c_t^\dagger\int_0^t (r^{-1}(c(s+u,X_u^{s,x'})-c(s+u,X_u^{s,x}))-\nabla c(s+u,X_u^{s,x})\cdot X_{u(\kappa)}^{s,x})du\bigg\|_{L^{k_2'}(\mathbb{P})} \\
&\quad \le \bigg\|\int_0^t\Big(c_u'\cdot r^{-1}(X_u^{s,x'}-X_u^{s,x})-\nabla c(s+u,X_u^{s,x})\cdot X_{u(\kappa)}^{s,x}\Big)du\bigg\|_{L^{k_2'}(\mathbb{P})} \\
&\quad \le \bigg\|\int_0^t c_u'\cdot(r^{-1}(X_u^{s,x'}-X_u^{s,x})-X_{u(\kappa)}^{s,x})du\bigg\|_{L^{k_2'}(\mathbb{P})} \\
&\qquad + \bigg\|\int_0^t (c_u'-\nabla c(s+u,X_u^{s,x}))\cdot X_{u(\kappa)}^{s,x}du\bigg\|_{L^{k_2'}(\mathbb{P})}.
\end{aligned}
\tag{4.4.8}
$$

For the first term of the right-hand side of (4.4.8), by Jensen's inequality, Theorem 2.4 in [101], setting $K > 2k_2'$ and our assumption about the derivatives of c, we have

$$
\begin{aligned}
&\mathbb{E}\left|\int_0^t c_u' \cdot \left(\frac{X_u^{s,x'} - X_u^{s,x}}{r} - X_{u(\kappa)}^{s,x}\right)du\right|^{k_2'} \\
&\leq T^{k_2'-1}\mathbb{E}\int_0^{T-s}\left|c_u' \cdot \left(\frac{X_u^{s,x'} - X_u^{s,x}}{r} - X_{u(\kappa)}^{s,x}\right)\right|^{k_2'} du \\
&\leq T^{k_2'-1}\left(\mathbb{E}\int_0^{T-s}|c_u'|^{2k_2'}du\right)^{\frac{1}{2}}\left(\mathbb{E}\int_0^{T-s}\left|\frac{X_u^{s,x+r\kappa} - X_u^{s,x}}{r} - X_{u(\kappa)}^{s,x}\right|^{2k_2'}du\right)^{\frac{1}{2}} \\
&\leq C(1 + V^{s,T}(0,\tilde{x}(x')) + V^{s,T}(0,\tilde{x}(x)))^{\frac{1}{2}} \\
&\quad \cdot \left(\mathbb{E}\sup_{0\leq u\leq T-s}\left|\frac{X_u^{s,x+r\kappa} - X_u^{s,x}}{r} - X_{u(\kappa)}^{s,x}\right|^{2k_2'}\right)^{\frac{1}{2}} \qquad (4.4.9)
\end{aligned}
$$

for C independent of t, which converges to 0 as $r \to 0$ by Lemma 4.3.2. For the second term on the right-hand side of (4.4.8), it holds that

$$
\begin{aligned}
&\mathbb{E}\left|\int_0^t (c_u' - \nabla c(s+u, X_u^{s,x})) \cdot X_{u(\kappa)}^{s,x}du\right|^{k_2'} \\
&\leq C\left(\int_0^{T-s}\mathbb{E}|c_u' - \nabla c(s+u, X_u^{s,x})|^{2k_2'}du\right)^{\frac{1}{2}}\left(\mathbb{E}\int_0^{T-s}|X_{u(\kappa)}^{s,x}|^{2k_2'}du\right)^{\frac{1}{2}}. \qquad (4.4.10)
\end{aligned}
$$

The last factor in the right-hand side of (4.4.10) is uniformly bounded in r by Lemma 4.3.2 and the first factor converges to 0 as $r \to 0$ by Lemma 4.4.3.

Putting together the above in (4.4.5) gives that $\mathbb{E}g(X_t^{s,x})e^{\int_0^t c(s+u,X_u^{s,x})du}$ is directionally differentiable in x. For the other term in (4.4.7), it suffices to check that after integrating the inequality (4.4.5) in t from 0 to $T-s$, the same convergences hold as $r \to 0$. This is true for the first three term on the right-hand side of (4.4.5) by the same reasoning as before. It is true for the right-hand side of (4.4.6) by dominated (in t) convergence, since the right-hand sides of (4.4.7), (4.4.9) and (4.4.10) are uniformly bounded in $t \in [0, T-s]$ and $r \in [0, \epsilon]$ for some $\epsilon > 0$. By induction and largely the same arguments as above, higher order directional derivatives in x of $\mathbb{E}\hat{h}(t,x)\hat{c}(t,x)$ exist and they are sums of expressions of the form

$$
\begin{aligned}
&\mathbb{E}\bigg[\partial^{\beta_1}h(s+t, X_t^{s,x})\hat{c}(t,x)\bigg(\prod_{\beta_2\in\hat{I}_2}(\partial^{(\beta_2)}X_t^{s,x})_{j_{\beta_2}}\bigg) \qquad (4.4.11) \\
&\quad \cdot \prod_{\beta_3\in\hat{I}_3}\int_0^t \partial^{\beta_3}c(s+u, X_u^{s,x})\prod_{\beta_4\in\hat{I}_{\beta_3}}(\partial^{(\beta_4)}X_u^{s,x})_{j_{\beta_4}}du\bigg],
\end{aligned}
$$

where $h \in \{f, g\}$, β_1 is a multiindex with $0 \leq |\beta_1| \leq p$, $\hat{I}_2, \hat{I}_3, \hat{I}_{\beta_3}$ are some finite sets of multiindices each with absolute value less than or equal to p and $j_{\beta_2}, j_{\beta_4} \in \{1, \ldots, n\}$. A fully detailed argument for this is omitted.

For differentiability of the expectation of (4.4.2) in x, note that Theorem 1.2 in [114] may be applied on (4.3.5) due to $\nabla b(s+t, X_t^{s,x}) \leq C(1 + \log V(t, \bar{x}(x))) \leq C(1 + V(t, \bar{x}(x)))$ (by Assumption 7 and the same for σ) and Lemma 4.3.1, so that the derivatives in probability $X_{t(\kappa)}$ are unique solutions to (4.3.5) for the initial condition κ. Therefore the first directional derivatives from (4.4.5) indeed form a linear map. The same arguments apply for expressions of the form (4.4.11) that are directionally differentiable, where additionally Assumption 8, Lemma 4.3.2 and Theorem 4.3.3 are to be used to control $K_t(1)$ from Theorem 1.2 in [114]. Next, we show continuity in x of expressions of the form (4.4.11) (for multiindices with absolute values bounded by p). Note first that $\mathbb{P}(\sup_{0 \leq u \leq T-s} |\partial^\beta X_u^{s,x'} - \partial^\beta X_u^{s,x}| \leq \epsilon) \leq \mathbb{P}(|\partial^\beta X_t^{s,x'} - \partial^\beta X_t^{s,x}| \leq \epsilon)$, therefore $\partial^\beta X_t^{s,x}$ is continuous in probability w.r.t. to x by Theorem 4.10 in [114]. Consequently the product w.r.t. β_2 in (4.4.11) and $\partial^{\beta_1} h(s+t, X_t^{s,x})$ are sequentially continuous in probability by Theorem 20.5 in [17]. Lemma 4.4.3 and continuous mapping theorem yield that $\hat{c}(t,x)$ is continuous in probability w.r.t. x. For the remaining factors in (4.4.11), for $1 < k < K$, we have

$$
\begin{aligned}
&\int_0^t \left| \partial^{\beta_3} c(s+u, X_u^{s,x'}) \prod_{\beta_4 \in \hat{I}_{\beta_3}} (\partial^{(\beta_4)} X_u^{s,x'})_{j_{\beta_4}} - \partial^{\beta_3} c(s+u, X_u^{s,x}) \prod_{\beta_4 \in \hat{I}_{\beta_3}} (\partial^{(\beta_4)} X_u^{s,x})_{j_{\beta_4}} \right| du \\
&\quad \leq \int_0^{T-s} \left| \partial^{\beta_3} (c(s+u, X_u^{s,x'}) - c(s+u, X_u^{s,x})) \prod_{\beta_4 \in \hat{I}_{\beta_3}} (\partial^{(\beta_4)} X_u^{s,x'})_{j_{\beta_4}} \right| du \\
&\qquad + \int_0^{T-s} \left| \partial^{\beta_3} c(s+u, X_u^{s,x}) \prod_{\beta_4 \in \hat{I}_{\beta_3}} (\partial^{(\beta_4)} (X_u^{s,x'} - X_u^{s,x}))_{j_{\beta_4}} \right| du \\
&\quad \leq C \int_0^{T-s} \left| \partial^{\beta_3} (c(s+u, X_u^{s,x'}) - c(s+u, X_u^{s,x})) \right| du \prod_{\beta_4 \in \hat{I}_{\beta_3}} \sup_{0 \leq u \leq T-s} \left| \partial^{(\beta_4)} X_u^{s,x'} \right| \\
&\qquad + C \int_0^{T-s} |\partial^{\beta_3} c(s+u, X_u^{s,x})| du \prod_{\beta_4 \in \hat{I}_{\beta_3}} \sup_{0 \leq u \leq T-s} \left| \partial^{(\beta_4)} X_u^{s,x'} - \partial^{(\beta_4)} X_u^{s,x} \right|
\end{aligned}
$$

By Hölder's inequality, Lemma 4.3.2 and Theorem 4.3.3, the first term on the right-hand side converges to zero in mean, therefore to zero in probability, as $x' \to x$. By Theorem 4.10 in [114] (and continuous mapping theorem), the second term on the right-hand side also converges to zero in probability. Therefore the left-hand side converges to zero in probability. By continuous mapping theorem, the term inside the square bracket in (4.4.11) is sequentially continuous in probability. Consequently, by (4.3.12) with $k_1 =$

1, $k = \frac{1+k_2}{2}$, $S = |J(x') - J(x)|$, where $J(x)$ is equal to the term inside the square brackets in (4.4.11), together with Hölder's inequality, inequality (4.4.1), our assumption on the derivatives of c with a large enough K, Theorem 2.4 in [101], Lemma 4.3.2 and Theorem 4.3.3, expectations of the form (4.4.11) are continuous functions w.r.t. x and so are their integrals in t by dominated convergence, which concludes the proof for (i).
Using the same results and denoting the expression (4.4.11) by $\hat{u}$, it holds that

$$\hat{u} \leq C(1+V^{s,T}(0,\tilde{x}(x)))^{\frac{1}{k_2}} \left(\mathbb{E}\left[\sup_{0\leq u\leq T-s} |\partial^{(\beta_2)} X_u^{s,x}|^{2k_2'}\right]\right)^{\frac{1}{2k_2'}}$$

$$\cdot \prod_{(\beta_3,\beta_4)\in\hat{I}} (1+V^{s,T}(0,\tilde{x}(x)))^{\frac{1}{c_{\beta_3}}} \left(\mathbb{E}\left[\sup_{0\leq u\leq T-s} |\partial^{(\beta_4)} X_u^{s,x}|^{c_{\beta_4}}\right]\right)^{\frac{1}{c_{\beta_4}}}$$

for some $c_{\beta_3}, c_{\beta_4} > 0$, $\beta_3, \beta_4 \in \hat{I}$ and in particular for some constant C independent of t. The proof for (ii) then concludes by Theorem 4.3.3.
Assertion (iii) then follows by Theorem 3.5(iii) in [114], Lemma 4.8.2 and by noting that C above is independent of s given that the Lyapunov functions are local in s.

□

4.4.2 Twice spatially differentiable solutions

In this section, we prove that the expectation of (4.4.2) with $t = T-s$ solves Kolmogorov's equation by the approach in [114]. The main ingredient beside differentiability of the associated semigroups, given in Theorem 4.4.2, is that the SDE can be approximated in probability by an Euler-type approximation locally uniformly in initial time and space, which is given in Lemma 4.4.4. Throughout this section, we assume $O = \mathbb{R}^n$.

Lemma 4.4.4. *Suppose for any $T > 0$, there exists a family of functions $(V^{s,T})_{s\in[0,T]}$ such that Assumption 9 holds. For $I = \{t_k\}_{k\in\mathbb{N}_0} \subset [0,\infty)$ with $t_0 = 0$, $t_{k+1} \geq t_k$, $k \in \mathbb{N}$, $t_k \to \infty$ as $k \to \infty$, $\sup_{k\geq 0} t_{k+1} - t_k < \infty$, $s \in [0,\infty)$, $x \in \mathbb{R}^n$, let $X_t^{s,x}(I)$ denote the Euler approximation given by $X_0^{s,x}(I) = x$ and*

$$X_t^{s,x}(I) = X_{t_k}^{s,x}(I) + \int_{t_k}^t b(s+r, X_{t_k}^{s,x}(I))dr + \int_{t_k}^t \sigma(s+r, X_{t_k}^{s,x}(I))dW_r, \qquad (4.4.12)$$

on $t \in [t_k, t_{k+1}]$, $k \in \mathbb{N}$. For any $R', T' \geq 0$, $\epsilon > 0$, it holds that

$$\sup_{s\in[0,T']} \sup_{|x|\leq R'} \mathbb{P}\left[\sup_{t\in[0,T']} |X_t^{s,x} - X_t^{s,x}(I)| \geq \epsilon\right] \to 0$$

as $\sup_{k\geq 0} t_{k+1} - t_k \to 0$.

Proof. We extend the proof of Theorem 1 in [164] to obtain convergence that is uniform with respect to $s \in [0,T]$ and $x \in B_R$. Fix the numbers $R', T' \geq 0$. For $k \in \mathbb{N}$, let $\varphi_k : \mathbb{R}^n \to [0,\infty)$ be smooth cutoff functions satisfying $\varphi_k(x) = 1$ for $x \in B_k$, $\varphi_k(x) = 0$ for $x \in \mathbb{R}^n \setminus B_{k+1}$ and let $b^{(k)} : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma^{(k)} : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ be given by $b^{(k)} = b\varphi_k$ and $\sigma^{(k)} = \sigma\varphi_k$. Let $Y_t^{s,x,k}(I)$ be the unique solutions to the corresponding SDE with drift $b^{(k)}$ and diffusion coefficient $\sigma^{(k)}$. The corresponding Euler approximation is given by (4.4.12) with $Y_0^{s,x,k} = Y_0^{s,x,k}(I) = x$. Fix w.l.o.g. $0 < \epsilon \leq 1$. In the same way as in the proof of Theorem 1 in [164], one obtains that for any $s \in [0,T']$, $x \in \mathbb{R}^n$ and $k \geq R' + 1$,

$$\mathbb{P}\left(\sup_{0\leq t\leq T'} |X_t^{s,x} - X_t^{s,x}(I)| > \epsilon\right) \leq \mathbb{P}\left(\sup_{0\leq t\leq T'} \left|Y_t^{s,x,k} - Y_t^{s,x,k}(I)\right| > \epsilon\right) + \mathbb{P}(\tau_{k-1} \leq T'),$$

where $\tau_{k-1} = \inf\{t \geq 0 : |X_t^{s,x}| > k-1\}$. By Markov's inequality, Theorem 2.4 in [101] and Assumption 9(iv), it holds that

$$\begin{aligned}
&\mathbb{P}(\tau_{k-1} \leq T') \inf_{t\in[s,s+T'],|y|=k-1} \hat{G}(t,y) \\
&\quad \leq \mathbb{E}[\hat{G}(s + (\tau_{k-1} \wedge T'), X^{s,x}_{\tau_{k-1}\wedge T'})] \\
&\quad \leq \mathbb{E}[V^{s,T'}(\tau_{k-1} \wedge T', \tilde{x}(x))] \\
&\quad \leq \left\| e^{\int_0^{(\tau_{k-1}\wedge T')} \alpha_u^{s,T'} du} \right\|_{L^{\frac{p^{s,T'}}{p^{s,T'}-1}}(\mathbb{P})} \left(V_0(s,\tilde{x}(x)) + \int_0^{T'} \left\| \frac{\mathbb{1}_{[0,\tau_{k-1}\wedge T')}(v)\beta_v^{s,T'}}{e^{\int_0^v \alpha_u^{s,T'} du}} \right\|_{L^{p^{s,T'}}(\mathbb{P})} dv \right).
\end{aligned}$$

For any $0 < \epsilon' < 1$, by the assumption that $V^{s,T'}$ is local in s and continuity of V_0, there exists k^* such that $\mathbb{P}(\tau_{k^*-1} \leq T') \leq \frac{\epsilon'}{2}$ for all $s \in [0,T']$ and $x \in B_{R'}$. In addition, for any $R > 0$, it holds that

$$\begin{aligned}
&2\langle x-y, b^{(k^*)}(t,x) - b^{(k^*)}(t,y)\rangle + \|\sigma^{(k^*)}(t,x) - \sigma^{(k^*)}(t,y)\|^2 \\
&\quad \leq 2\langle x-y, b(t,x) - b(t,y)\rangle \varphi_{k^*}(x) + 2|b(t,y)||x-y||\varphi_{k^*}(x) - \varphi_{k^*}(y)| \\
&\qquad + \|\sigma(t,x) - \sigma(t,y)\|^2 \varphi_{k^*}(x)^2 + \|\sigma(t,y)\|^2 |\varphi_{k^*}(x) - \varphi_{k^*}(y)|^2 \\
&\quad \leq (K_t(R) + C \sup_{y'\in B_R} (|b(t,y')| + \|\sigma(t,y')\|^2))|x-y|^2
\end{aligned}$$

for all $x, y \in B_R$ and

$$2\langle x, b^{(k^*)}(t,x)\rangle + \|\sigma^{(k^*)}(t,x)\|^2 \leq 2(1+|x|) \sup_{x'\in B_{k^*+1}} (|b(t,x')| + \|\sigma(t,x')\|^2)$$

for all $x \in \mathbb{R}^n$. Therefore Corollary 5.4 in [114] can be applied to obtain

$$\sup_{s\in[0,T']} \sup_{x\in B_R} \mathbb{P}\left(\sup_{0\le t\le T'} \left| Y_t^{s,x,k^*} - Y_t^{s,x,k^*}(I) \right| > \epsilon \right) \to 0$$

as $\sup_{k\ge 0} t_{k+1} - t_k \to 0$, which concludes the proof. $\square$

Theorem 4.4.5. *Let all of the assumptions in Theorem 4.4.2 hold. In particular, let c have Lyapunov derivatives up to order (p, k') for any $1 < k' < K$, let the mappings $\hat{x}_k$ be independent of s, $(\hat{V}_k^{s,T})_{s\in[0,T]}$ be local in s for any k, multiindex α with $0 \le |\alpha| \le p$ and let $p \ge 2$. For $v : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ given by*

$$v(t,x) = \mathbb{E}u(t, T-t, x), \tag{4.4.13}$$

with u as in (4.4.2)*, the equation*

$$\partial_t v + a : D^2 v + b \cdot \nabla v - cv + f = 0 \tag{4.4.14}$$

holds almost everywhere in $(0,T) \times \mathbb{R}^n$.

Proof. Theorem 4.4.2, Theorem 3.6 in [114] applied on the SDE (4.1.1) appended by (4.8.2) and Lemma 4.8.2 yield $(\partial_t v + b \cdot \nabla v + a : D^2 v - cv + f)e^{-x'} = 0$ almost everywhere. $\square$

Note the assumptions in Theorems 4.4.2 and 4.4.5 remain strictly weaker than those in [114, Lemma 5.10], since Lyapunov functions that are positive polynomials can easily be conjured under the global Lipschitz conditions there.

4.4.3 Distributional solutions under maximal dissipativity

This section complements our result about Kolmogorov equations by considering a case where maximal dissipativity of the closure of the generator acting on C_c^∞ is known. It is shown that if there exists an associated semigroup and the coefficients b and σ are regular enough, this is sufficient for a solution of the backward Kolmogorov equation in the distributional sense. In this section we do not assume our conditions about the Lipschitz constants or the higher derivatives of the coefficients. Below, $\mathcal{L}$ is used to denote the differential operator

$$\mathcal{L} = b \cdot \nabla + a : D^2$$

defined on C_c^∞.

Proposition 4.4.6. *Assume $O = \mathbb{R}^n$, a and b are independent of ω, t, they admit distributional derivatives of order two and one respectively and that μ is a probability measure* *on $\mathbb{R}^n$ absolutely continuous with respect to the Lebesgue measure with density ρ satisfying, for some $p, q \in [1, \infty]$ with $p^{-1} + q^{-1} = 1$,*

- *$\rho^{-1}, \rho^{-1}\partial_i^k \partial_j^l a_{ij}, \rho^{-1}\partial_i^k b_i \in L^p_{loc}(\mu)$, for $i, j \in \{1, \dots, n\}$, $k, l \in \{0, 1\}$, where $\rho^{-1} := 0$ whenever $\rho = 0$,*
- *the closure $\bar{L}$ of $\mathcal{L}$ in $L^q(\mu)$ generates a strongly continuous semigroup $\{T_t\}_{t \geq 0}$ on $L^q(\mu)$,*

then for any $g \in \mathcal{D}(\bar{L})$ and $\eta \in C_c^\infty((0, T) \times \mathbb{R}^n)$,

$$\int_0^T \int_{\mathbb{R}^n} T_t g \left[\frac{\partial \eta}{\partial t} + \sum_{ij} \partial_i \partial_j (a_{ij}\eta) - \sum_i \partial_i (b_i \eta) \right] dx dt = 0. \tag{4.4.15}$$

In the case when $\det(a) > 0$ and a, b are regular enough, Sections 3.4, 3.5, 5.2 in [20] and Section 8.1 in [128] provide results sufficient for the assumptions in Proposition 4.4.6. Otherwise when $\det(a) > 0$ does not hold everywhere, such statements are less generally available, but hold for example in the settings of [43][1] and [11]; note the assumed regularity on a, b are more than what's required in these works but are necessary for the formulation of (4.4.15).
Stationarity of μ is not required for the proof of Proposition 4.4.6, but it is the case for the assumptions to be satisfied in the above references. In addition, $\{T_t\}_{t\geq 0}$ is not specified in terms of an expectation as in (4.4.13); for this, there must be some stochastic process X_t associated to $\{T_t\}_{t\geq 0}$, which solves (4.2.1). In [43] and [11] (for example), where such a process is given, the statement in Proposition 4.4.6 does not amount to a corollary of (the proof of) Proposition 4.18 in [85] because the probability measures in the associated Hunt process are not necessarily shared amongst different initial points.

Proof of Proposition 4.4.6. By mean value theorem, there exists constant $k_\eta > 0$ and compact set $K_\eta \subset (0, T) \times \mathbb{R}^n$ such that $\left| T_t g \frac{\eta(t+s, \cdot) - \eta(t, \cdot)}{s} \right| \leq |T_t g| k_\eta \mathbb{1}_{K_\eta}$ for all s, therefore by the dominated convergence theorem,

$$\int_{T_\epsilon} \int_{\mathbb{R}^n} T_t g \partial_t \eta dx dt = -\lim_{s \to 0} \int_0^T \int_{\mathbb{R}^n} T_t g \frac{\eta(t-s, x) - \eta(t, x)}{s} dx dt.$$

By the inequality above, strong continuity of T_t and the assumption on ρ^{-1}, the expressions above make sense and the integral under the limit has the same limit as the left-hand

[1] The state space here is slightly different to $\mathbb{R}^n$, but the statement and proof of Proposition 4.4.6 can be modified accordingly.

side of

$$\int_0^{T+s}\int_{\mathbb{R}^n} T_t g \frac{\eta(t-s,x)-\eta(t,x)}{s} dxdt = \int_0^T \int_{\mathbb{R}^n} \frac{T_{t+s}g - T_t g}{s} \eta dxdt,$$

which can be dealt with by considering

$$\begin{aligned}
&\left|\int_0^T \int_{\mathbb{R}^n} \left(\frac{T_{t+s}g - T_t g}{s} - \bar{L}T_t g\right)\eta dxdt\right| \\
&\quad \leq \int_0^T \left\|\frac{T_{t+s}g - T_t g}{s} - \bar{L}T_t g\right\|_{L^q(\mu)} \|\eta(t,\cdot)\rho^{-1}\|_{L^p(\mu)} dt \\
&\quad \leq k_\mu \int_0^T \left\|\frac{T_{t+s}g - T_t g}{s} - \bar{L}T_t g\right\|_{L^q(\mu)} dt \qquad (4.4.16)
\end{aligned}$$

for some constant $k_\mu > 0$. The right hand side of (4.4.16) is converging to zero as $s \to 0$ by dominated convergence theorem with constant dominating function since

$$\begin{aligned}
\left\|\frac{T_{t+s}g - T_t g}{s} - \bar{L}T_t g\right\|_{L^q(\mu)} &\leq \left\|T_t\left(\frac{T_s g - g}{s}\right)\right\|_{L^q(\mu)} + \|T_t \bar{L} g\|_{L^q(\mu)} \\
&\leq M e^{\omega T}(1 + 2\|\bar{L}g\|_{L^q(\mu)})
\end{aligned}$$

for all $s \in (0,S)$, $t \in [0,T]$, some $M, S > 0$ and $\omega > 0$ by Proposition 1.1 in [63]. Therefore,

$$\int_0^T \int_{\mathbb{R}^n} T_t g \partial_t \eta dxdt = -\int_0^T \int_{\mathbb{R}^n} \bar{L}T_t g \eta dxdt.$$

By assumption, for each $t \in (0,T)$, there exist a sequence $(g_k^t)_{k\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^n)$ with $g_k^t \to T_t g$ and $\mathcal{L}g_k^t = \bar{L}g_k^t \to \bar{L}T_t g$ in $L^q(\mu)$ as $k \to \infty$. Since $\rho^{-1} \in L^p_{loc}(\mu)$, we have for every t,

$$\begin{aligned}
\left|\int_{\mathbb{R}^n} (\bar{L}T_t g - \mathcal{L}g_k^t)\eta dx\right| &\leq \|\bar{L}T_t g - \mathcal{L}g_k^t\|_{L^q(\mu)} \|\rho^{-1}\eta\|_{L^p(\mu)} \\
\left|\int_{\mathbb{R}^n} (T_t g - g_k^t)\mathcal{L}^\top \eta dx\right| &\leq \|T_t g - g_k^t\|_{L^q(\mu)} \|\rho^{-1}\mathcal{L}^\top \eta\|_{L^p(\mu)},
\end{aligned}$$

where $\mathcal{L}^\top$ denotes the $L^2(\mathbb{R}^n)$-adjoint of $\mathcal{L}$, which concludes the proof. □

4.5 Alternative assumptions for time-independent, nonrandom coefficients

In the following, we restrict to the case where b and σ are nonrandom and time-independent, so that we may use Theorem V.39 in [165] in order to rid the need for bounds on function values on line segments in terms of the endpoint values. In doing so, more local conditions are obtained in place of (4.3.1), (4.3.2), (4.3.13) and (4.4.1).

Lemma 4.5.1. *Let $p \in \mathbb{N}$, b, σ be independent of ω, t and suppose they are continuously differentiable up to order p with locally Lipschitz derivatives. For every $s \in [0,T]$, there exists $\Omega \times \Delta_T \times \mathbb{R}^n \ni (\omega, t, x) \mapsto \hat{X}_t^x \in \mathbb{R}^n$ that is for $\mathbb{P}$-a.a. $\omega \in \Omega$ continuously differentiable in x up to order p and indistinguishable from the corresponding derivatives in probability of $X_\cdot^x$.*

Proof. By Theorem V.38 and V.39 in [165], continuously differentiable $\hat{X}_\cdot^x$ up to order p exists. Moreover, it satisfies (4.2.1) and $\hat{X}_\cdot^x$ is indistinguishable from $X_\cdot^x$. The partial derivatives of $\hat{X}_\cdot^x$ satisfy the systems given by formal differentiation of (4.2.1). On the other hand, derivatives in probability of X_t^x as in [114, Theorem 4.10] and Theorem 3.3 above satisfy the same system. Therefore by uniqueness in the aforementioned references[2], it holds that $\partial^\alpha \hat{X}_\cdot^x$ are the unique solutions to their respective systems for all time and are therefore indistinguishable from the corresponding derivatives in probability $\partial^{(\kappa_\alpha)} X_\cdot^x$ for every $s \in [0,T]$ and multiindex α with $0 \le |\alpha| \le p$. □

Theorem 4.5.2 (Alternative assumptions to Lemma 4.3.2 and Theorems 4.3.3, 4.4.2 and 4.4.5)**.** *Let b and σ be independent of ω, t and let $O = \mathbb{R}^n$. The following statements hold.*

(i) Lemma 4.3.2 continues to hold with $W(x, r\kappa) = (1 + \int_0^1 V(0, \bar{x}(x + \lambda r\kappa))d\lambda)$ if

- *the coefficients b and σ admit locally Lipschitz first derivatives and*
- *in Assumption 7, the inequalities (4.3.1), (4.3.2) are replaced by*

$$\sum_i |\partial_i b(x)| + \|\partial_i \sigma(x)\|^2 \le G(t,x). \tag{4.5.1}$$

(ii) Theorem 4.3.3 continues to hold with (4.3.17) replaced by

$$q(y,y') = q_0\bigg(\int_0^1 V(0, \bar{x}(\lambda y + (1-\lambda)y'))d\lambda), \int_0^1 \hat{V}_{l_1}^{s,T}(0, \hat{x}_{l_1}(\lambda y + (1-\lambda)y'))d\lambda,$$
$$V(0, \bar{x}(y)), (\hat{V}_{l_i}^{s,T}(0, \hat{x}_{l_i}(y)))_{i \in \{2,\dots,i^*\}}, (\hat{V}_{l_i}^{s,T}(0, \hat{x}_{l_i}(y')))_{i \in \{2,\dots,i^*\}}\bigg)$$

if

- *the coefficients b and σ admit locally Lipschitz second derivatives,*

[2] Alternatively, since these systems have terms on right-hand sides that are continuous functions of the partial derivatives and are in particular at most linear in the highest order derivative (see the beginning of proof for Theorem 4.3.3), uniqueness holds by continuity of X_t^x in t, (4.3.7) in Lemma 4.3.2, induction in the number of derivatives and Theorem 1.2 in [114] with $K_t(R) = K_t(1)$ constant in t.

- *Assumption 7 is replaced as above and*
- *in Assumption 8, inequality* (4.3.13) *is replaced by*

$$|\partial^\alpha b(X_t^x)| + \|\partial^\alpha \sigma(X_t^{x'})\|^2 \le M'(1 + \hat{V}_k^{s,T}(t, \hat{x}_k(y)))^{\frac{1}{k}}.$$

(iii) Theorems 4.4.2 and 4.4.5 continue to hold if

- *the second derivatives of b and σ are locally Lipschitz,*
- *Assumption 7 and 8 are replaced as above and*
- *in Definition 4.4.1, inequality* (4.4.1) *is replaced by*

$$|\partial^\alpha h(X_t^x)| \le N(1 + V^{s,T}(t, \tilde{x}(x)))^{\frac{1}{k}}.$$

Proof. The proof strategies follow largely in the same way as in the previous proofs, the differences are specified in the following using the same notation as before. For (i), note first (4.3.7) follows unperturbed. By Lemma 4.5.1, classical derivatives are indistinguishable from derivatives in probability and we use the properties of both without changing the notation in the following. In place of (4.3.9), it holds that

$$dX_{t(\kappa)}^{(r)} = r\int_0^1 (X_{t(\kappa)}^{x+\lambda r\kappa} \cdot \nabla) b(X_t^{x+\lambda r\kappa}) d\lambda dt + r\int_0^1 (X_{t(\kappa)}^{x+\lambda r\kappa} \cdot \nabla)\sigma(X_t^{x+\lambda r\kappa}) d\lambda dW_t.$$

Note that since for every t and almost all ω, the functions X_t^x, $X_{t(\kappa)}^x$ are continuous in x, the integrands on the right-hand side are $\mathcal{B}([0,T]) \otimes \mathcal{F} \otimes \mathcal{B}([0,1])$-measurable by Lemma 4.51 in [2] and the integrals (in λ) themselves are adapted. For any $\hat{k} \ge 1$, by (4.5.1), the coefficients satisfy

$$\begin{aligned}
&2rX_{t(\kappa)}^{(r)} \cdot \int_0^1 (X_{t(\kappa)}^{x+\lambda r\kappa} \cdot \nabla) b(X_t^{x+\lambda r\kappa}) d\lambda \\
&\quad + (2\hat{k} - 1)\left\| r\int_0^1 (X_{t(\kappa)}^{x+\lambda r\kappa} \cdot \nabla)\sigma(X_t^{x+\lambda r\kappa}) d\lambda \right\|^2 \\
&\qquad \le \left| X_{t(\kappa)}^{(r)} \right|^2 + 2\hat{k}r^2 \int_0^1 \left| X_{t(\kappa)}^{x+\lambda r\kappa} \right|^2 (2G(t, X_t^{x+\lambda r\kappa}))^2 d\lambda.
\end{aligned}$$

Consequently, Theorem 2.4 in [101] (in place of Corollary 2.5 in the proof of Lemma 4.3.2) can be applied with

$$a_t = r\int_0^1 (X_{t(\kappa)}^{x+\lambda r\kappa}\cdot\nabla)b(X_{t(\kappa)}^{x+\lambda r\kappa})d\lambda,\ b_t = r\int_0^1 (X_{t(\kappa)}^{x+\lambda r\kappa}\cdot\nabla)\sigma(X_{t(\kappa)}^{x+\lambda r\kappa})d\lambda,$$
$$p = 2k\vee 2,\ \alpha_t = 1,\ \beta_t = 4(k\vee 1)r^2\int_0^1 \left|X_{t(\kappa)}^{x+\lambda r\kappa}\right|^2 G(t, X_t^{x+\lambda r\kappa})^2 d\lambda,$$
$$q_1 = \frac{k}{2},\ q_2 = \left(\frac{2}{k} - \frac{1}{k\vee 1}\right)^{-1},\ q_3 = k\vee 1,\ V(x) = |x|^2,$$

to obtain

$$\begin{aligned}
&\mathbb{E}\sup_{0\le u\le t}|X_u^x|^k \\
&\quad\le Ce^{\frac{kt}{2}}r^k\left(\mathbb{E}\left[1+4(k\vee 1)\int_0^t e^{-s}\int_0^1\left|X_{s(\kappa)}^{x+\lambda r\kappa}\right|^2 G(s, X_s^{x+\lambda r\kappa})^2 d\lambda ds\right]^{k\vee 1}\right)^{\frac{k}{2(k\vee 1)}} \\
&\quad\le Ce^{\frac{kt}{2}}r^k\left(\mathbb{E}\left[1+\int_0^1\sup_{0\le u\le t}\left|X_{u(\kappa)}^{x+\lambda r\kappa}\right|^{2k\vee 2}\int_0^t G(s, X_s^{x+\lambda r\kappa})^{2k\vee 2} ds d\lambda\right]\right)^{\frac{k}{2(k\vee 1)}} \\
&\quad\le Ce^{\frac{kt}{2}}r^k\left(1+\left(\int_0^1\mathbb{E}\sup_{0\le u\le t}\left|X_{u(\kappa)}^{x+\lambda r\kappa}\right|^{4k\vee 4}d\lambda\right)^{\frac{1}{2}}\right. \\
&\qquad\left.\cdot\left(\mathbb{E}\left[\int_0^1\int_0^t G(s, X_s^{x+\lambda r\kappa})^{2k\vee 2}ds d\lambda\right]^2\right)^{\frac{1}{2}}\right)^{\frac{k}{2(k\vee 1)}}.
\end{aligned}$$

By (4.3.7), the first expectation on the right-hand side has the bound

$$\mathbb{E}\sup_{0\le u\le t}\left|X_{u(\kappa)}^{x+\lambda r\kappa}\right|^{4k\vee 4}d\lambda \le \int_0^1 \rho(1+V(0,\bar{x}(x+\lambda r\kappa)))d\lambda$$

and, by (4.3.3) and Lemma 4.3.1, the second expectation has the bound

$$\begin{aligned}
\mathbb{E}\left[\int_0^1\int_0^t G(u, X_u^{x+\lambda r\kappa})^{2k\vee 2}du d\lambda\right]^2 &\le \int_0^1 t\int_0^t \mathbb{E}G(u, X_u^{x+\lambda r\kappa})^{4k\vee 4}du d\lambda \\
&\le C\int_0^1 t\int_0^t \mathbb{E}(1+\log V(u,\bar{x}(x+\lambda r\kappa)))^{4k\vee 4}du d\lambda \\
&\le C\int_0^1 t\int_0^t \mathbb{E}(1+V(u,\bar{x}(x+\lambda r\kappa)))du d\lambda \\
&\le Ct\int_0^1 (V(0,\bar{x}(x+\lambda r\kappa))+1)d\lambda,
\end{aligned}$$

which concludes the proof of (i).

For (ii), the conclusions of Theorem 4.3.3 follow with differences that have already been addressed when dealing with (i), using that expressions of the form $h(s+u, X_u^{s,x'}) -$

$h(s+u, X_u^{s,x}) = \int_0^1 \nabla h(s+u, \lambda X_u^{s,x'} + (1-\lambda)X_u^{s,x}) \cdot (X_u^{s,x'} - X_u^{s,x})d\lambda$ may be replaced by $\int_0^1 \nabla h(t, X_u^{x+\lambda r\kappa}) \cdot r X_{u(\kappa)}^{x+\lambda r\kappa} d\lambda$ and therefore the arguments are not repeated.
For (iii), Lemma 4.4.3 can easily be modified using what has already been mentioned, so that Theorem 4.4.2(i) holds. Proofs for the other assertions of Theorem 4.4.2 and Theorem 4.4.5 follow unperturbed. □

4.6 Weak convergence rates for approximations under Lyapunov conditions

Here, the results in Section 4.3 are used with the exponential integrability property of stopped increment-tamed Euler-Maruyama schemes from [105] in order to establish weak convergence rates for SDEs with non-globally monotone coefficients. Classical proofs as in [112] establishing weak rates for the Euler-Maruyama scheme approximating (4.2.1) with globally Lipschitz coefficients require bounds on derivatives of the expectation (4.4.13), the Kolmogorov equation (4.4.14) and moment bounds on the discretisation. Although analogous requirements have mostly (beside continuous differentiability of (4.4.13) in t, which may be remedied for example by hypoellipticity in some cases) been shown to hold to an extent in the setting here, the Itô-Alekseev-Gröbner formula of [99] is used for a more direct proof, which uses moment estimates on derivative processes as the main prerequisites. Along the way, strong completeness (see e.g. [126] for a definition) of the derivative SDEs as in (4.3.5) (and its higher order analogues) are shown in Lemma 4.6.2 using a result of [44]. The same assertions as those in Lemma 4.6.2 up to order 2 have appeared recently in [100] under different assumptions. The approach here uses the results in [165] for continuous differentiability in initial condition as a starting point and consequently requires (at least at face value) the underlying space to be all of $\mathbb{R}^n$. Before the aforementioned strong completeness result, a local Hölder continuity in time result in the strong $L^p(\mathbb{P})$ sense for derivatives to our SDE is shown in Lemma 4.6.1. We begin by stating the numerical scheme and assumptions from [105] (amongst which is a Lyapunov-type condition) used for its exponential integrability. Assumptions about the relationship between the Lyapunov(-type) functions there and those in Assumptions 7, 8 are stated alongside, as well as the mild assumptions from [99]. Lemma 4.6.2 serves to verify the more serious assumptions in [99, Theorem 3.1] for use in proving the main Theorem 4.6.3.

Assumption 10. (i) The filtration $\mathcal{F}_t$ satisfies $\mathcal{F}_t = \sigma(\mathcal{F}_0 \cup \sigma(W_s : s \in [0,t]) \cup \{A \in \mathcal{F} : \mathbb{P}(A) = 0\})$ and that $\mathcal{F}_0$ and $\sigma(W_s : s \in [0,T])$ are independent. It holds that $O = \mathbb{R}^n$ and b, σ are independent of ω, t.

(ii) There exist $\gamma, \rho \geq 0$, $\gamma', c' > 0$, $\xi, c > 1$, $C \in \mathbb{R}$, $U \in C^2(\mathbb{R}^n, [0,\infty))$, $\bar{U} \in C(\mathbb{R}^n)$

such that $\bar{U} > C$, $U(x) \geq c'|1+x|^{\gamma'}$ and

$$
\begin{aligned}
&\sup_{\substack{\kappa_1,\dots,\kappa_j\in\mathbb{R}^n\setminus\{0\}:\\|\kappa_1|=\dots=|\kappa_j|=1}}\left|\sum_{i_1,\dots,i_j=1}^{n}\partial_{i_1}\dots\partial_{i_j}(U(x)-U(y))(\kappa_1)_{i_1}\dots(\kappa_j)_{i_j}\right| \\
&\qquad \leq c|x-y|\left(1+\sup_{\lambda\in[0,1]}|U(\lambda x+(1-\lambda)y)|\right)^{(1-\frac{j+1}{\xi})\vee 0}, \\
&|\partial^\alpha b(x)| + \|\partial^\alpha\sigma(x)\| + |\bar{U}(x)| \leq c(1+U(x))^\gamma, \\
&\frac{|\bar{U}(x)-\bar{U}(y)|}{|x-y|} \leq c(1+|U(x)|^\gamma + |U(y)|^\gamma), \\
&LU(x) + \frac{1}{2}\|\sigma^\top\nabla U(x)\|^2 + \bar{U}(x) \leq \rho U(x).
\end{aligned}
$$

for all $x, y \in \mathbb{R}^n$, $j \in \{0,1,2\}$ and multiindices α with $0 \leq |\alpha| \leq 2$.

(iii) For any $\theta \in \Theta := \{\theta = (t_0,\dots,t_{n^*}) : n^* \in \mathbb{N}, t_k \in [0,T], t_k < t_{k+1}, k \in \{1,\dots,n^*-1\}, t_0 = 0, t_{n^*} = T\}$, the function $Y_\cdot^\theta : \Omega\times[0,T] \to \mathbb{R}^n$ is an $\mathcal{F}_t$-adapted, $\mathbb{P}$-a.s. continuous process satisfying $\sup_{\theta\in\Theta}\mathbb{E}[e^{U(Y_0^\theta)}] < \infty$ and

$$
\begin{aligned}
Y_t^\theta = Y_{t_k}^\theta &+ \mathbb{1}_{\{y:|y|<\exp(|\log\sup_k t_{k+1}-t_k|^{\frac{1}{2}})\}}(Y_{t_k}^\theta) \\
&\cdot\left[\frac{b(Y_{t_k}^\theta)(t-t_k)+\sigma(Y_{t_k}^\theta)(W_t-W_{t_k})}{1+|b(Y_{t_k}^\theta)(t-t_k)+\sigma(Y_{t_k}^\theta)(W_t-W_{t_k})|^{q'}}\right]
\end{aligned}
$$

on $t \in [t_i, t_{i+1})$ for each $k \in \{0,\dots,n^*-1\}$, where $q' \geq 3$.

(iv) Assumptions 7 and 8 hold and $p \geq 3$. For any $V' \in \{V, \hat{V}_k^{s,T} : s \in [0,T], 2 \leq |\alpha| \leq p-1, k \geq 2\}$, there exist $0 < l^* \leq 1$, $\bar{n} \geq n$, $\bar{O} \subset \mathbb{R}^{\bar{n}}$ and $\hat{b}, \hat{\sigma}$ such that V' is a $(\tilde{b}_\cdot, \tilde{\sigma}_\cdot, C, 0, 1, V_0)$-Lyapunov function for some $\tilde{b}_\cdot$, $\tilde{\sigma}_\cdot$, V_0 with $\tilde{b}_t^y = \hat{b}(t, Y_t^y)$, $\tilde{\sigma}_t^y = \hat{\sigma}(t, Y_t^y)$ for processes Y_t^y satisfying (4.2.3) and $V_0 \in C^2([0,T],\bar{O})$ satisfies $\mathbb{P}$-a.s. that

$$
\begin{aligned}
&(\partial_t + L)V_0(t,y) \leq CV_0(t,x), && \lim_{|x'|\to\infty} V_0(t,x') = \infty, && (4.6.1)\\
&V'(0,\tilde{x}'(X_{s,t}^y))^{l^*} \leq C(1+V'(t-s,\tilde{x}'(y))), && V_0(0,\tilde{x}'(y))^{l^*} \leq C(1+e^{U(y)e^{-\rho T}})
\end{aligned}
$$

for all $s,t \in [0,T]$, $x \in \bar{O}$, $y \in \cup_{\theta\in\Theta}\mathrm{Range}(Y_\cdot^\theta)$, where $\tilde{x}' = \bar{x}$ if $V' = V$, $\tilde{x}' = \hat{x}_k$ otherwise, L is given by (4.2.2) with b, σ replaced by $\hat{b}, \hat{\sigma}$ and $X_{s,\cdot}^y$ is the solution to

$$
X_{s,t}^y = y + \int_s^t b(X_{s,u}^y)du + \int_s^t \sigma(X_{s,u}^y)dW_u. \qquad (4.6.2)
$$

Remark 4.6.1. By Theorem 3.5 in [111], the first part of Assumption 10(iv) implies that for all $s \in [0,T]$, $x \in \mathbb{R}^n$, there exists a unique up to distinguishability, $\mathcal{F}_t$-adapted, $\mathbb{P}$-a.s. continuous solution to (4.6.2) and for $t \in [s,T]$ it holds $\mathbb{P}$-a.s. that $X_{t,T}^{X_{s,t}^x} = X_{s,T}^x$. In (i), the assertions about $\mathcal{F}_t$ are from [99]. We set O to be the whole space and fix b and σ to be time-independent and nonrandom in order to use continuous differentiability in initial value from [165] and to use the exponential integrability results of [105]. Items (ii) and (iii) closely follow the assumptions in [105]. Here, of particular note is that q' is asserted to be greater than or equal to 3 rather than 1 in the denominator of the expression for Y_t^θ; this assumption is made in order to ensure well-behavedness of some higher order terms in the Itô-Alekseev-Gröbner expansion such that weak convergence rate of order 1 is attained. It is worth mentioning that the Lipschitz estimate on U with $j=0$ in (ii) easily gives that U is polynomially bounded, so that the set under the indicator function in (iii) indeed satisfies the assumptions in [105], as used in [103, 105]. The last assertions of item (iv) (and in general Assumption 10) are easily satisfied by all of the examples mentioned here; they collect properties of the Lyapunov-type function from (ii) required for our argument without requiring the Lyapunov functions to have V_0 be given by $e^{U(x)e^{-\rho t}+y}$ (see the proof of Corollary 3.3 in [101]).

In the following, for any $s \in [0,T]$, we extend the definition of any process Z_t defined on $[s,T]$ to $[0,T]$ by setting $Z_t = Z_s$ for $t \in [0,s)$.

Lemma 4.6.1. *Under Assumption 10, for any $k_1 > 2(n+1)$, $R > 0$, there exist constants $C > 0$, $n+1 < \nu_1 \leq k_1$ such that*

$$\mathbb{E}\sup_{u\in[s,t]} |\partial^{(\kappa)} X_{s,u}^x - \partial^{(\kappa)} X_{s,s}^x|^{k_1} < C|t-s|^{\nu_1}$$

for all $(s,t) \in \Delta_T$, $x \in B_R$, $\kappa \in \{(\kappa_i)_{1\leq i\leq p_0} : \kappa_i \in \mathbb{R}^n, |\kappa_i| = 1, 1 \leq i \leq p, p_0 \in \mathbb{N}_0 \cap [0,p]\}$.

Proof. By (4.3.15) in Theorem 4.3.3 (with a time shifted Wiener process and filtration) and using that $\partial^{(\kappa)} X_{s,s}^x = 0$ (for κ in the following set), the existence of such constants have already been shown for $\kappa \in \{(\kappa_i)_{1\leq i\leq p_0} : \kappa_i \in \mathbb{R}^n, |\kappa_i| = 1, 1 \leq i \leq p, p_0 \in \mathbb{N}_0 \cap [2,p]\}$. Using Assumption 10(ii), Corollary 2.5 and Corollary 3.3 both in [101], it holds that

$$\begin{aligned}
\mathbb{E}\sup_{u\in[s,t]} |X_{s,u}^x - x|^{k_1} &\leq Ce^{k_1(t-s)}\left(\int_0^{t-s}\left(\mathbb{E}\left[e^{U(X_{s,s+u}^x)e^{-\rho u}-2k_1 u}\right]\right)^{\frac{1}{k_1}} du\right)^{\frac{k_1}{2}} \\
&\leq Ce^{k_1(t-s)}\left(\int_0^{t-s} e^{\frac{U(x)}{k_1}} du\right)^{\frac{k_1}{2}} \\
&\leq C|t-s|^{\frac{k_1}{2}}
\end{aligned}$$

for all $(s,t) \in \Delta_T$, $x \in B_R$. Using instead Assumption 7, it holds that

$$
\begin{aligned}
&\mathbb{E} \sup_{u\in[s,t]} |\partial^{(\kappa_i)} X^x_{s,u} - \kappa_i|^{k_1} \\
&\leq C\Big(\mathbb{E}\Big[e^{\int_0^{t-s} \frac{1}{2k_1 T} \log V(u,\bar{x}(x))}\Big]^{2k_1}\Big)^{\frac{1}{2}} \bigg(\int_0^{t-s} (\mathbb{E}[\log V(u,\bar{x}(x)) + 1]^{2k_1})^{\frac{1}{k_1}} du\bigg)^{\frac{k_1}{2}} \\
&\leq C(1+V(0,\bar{x}(x)))\bigg(\int_0^{t-s} (\mathbb{E}V(u,\bar{x}(x)) + 1)^{\frac{1}{k_1}} du\bigg)^{\frac{k_1}{2}} \\
&\leq C(1+V(0,\bar{x}(x)))((t-s)(V(0,\bar{x}(x))+1)^{\frac{1}{k_1}})^{\frac{k_1}{2}} \\
&\leq C|t-s|^{\frac{k_1}{2}}
\end{aligned}
$$

for all $(s,t) \in \Delta_T$, $x \in B_R$, $\kappa_i \in \mathbb{R}^n$ with $|\kappa_i| = 1$. □

The following lemma shows that the assumptions of Theorem 3.1 in [99] hold under Assumption 10. Moreover, it is shown that the estimates therein hold uniformly with respect to the discretisation $\theta \in \Theta$.

Lemma 4.6.2. *Let Assumption 10 hold. There exists a function $\Omega \times \Delta_T \times \mathbb{R}^n \ni (\omega,(s,t),x) \mapsto \bar{X}^x_{s,t}(\omega) \in \mathbb{R}^n$ such that*

- *it holds $\mathbb{P}$-a.s. that for any $(s,t) \in \Delta_T$, $\mathbb{R}^n \ni x \mapsto \bar{X}^x_{s,t} \in \mathbb{R}^n$ is continuously differentiable in x up to order $p-1$ and the derivative $\Delta_T \times \mathbb{R}^n \ni ((s,t),x) \mapsto \partial^\alpha \bar{X}^x_{s,t} \in \mathbb{R}^n$ is continuous for all multiindices α with $0 \leq |\alpha| \leq p-1$,*
- *for any $s \in [0,T]$, $x \in \mathbb{R}^n$, the function $\partial^\alpha \bar{X}^x_{s,\cdot}$ is indistinguishable from $\partial^{(\kappa_\alpha)} X^x_{s,\cdot}$ for all multiindices α with $0 \leq |\alpha| \leq p-1$.*

Moreover, for any $p^\dagger > 0$, it holds that

$$
\sup_{0\leq|\alpha|\leq p-1} \sup_{\theta\in\Theta} \sup_{0\leq r\leq s\leq t\leq T} \mathbb{E}\Big[|b(\bar{X}^{Y^\theta_s}_{s,t})|^{p^\dagger} + \|\sigma(\bar{X}^{Y^\theta_s}_{s,t})\|^{p^\dagger} + |\partial^\alpha \bar{X}^{\bar{X}^{Y^\theta_r}_{r,s}}_{t,T}|^{p^\dagger}\Big] < \infty.
$$

Proof. By Lemma 4.5.1 (with time-shifted Wiener process and filtration), derivatives in probability $\partial^{(\kappa_\alpha)} X^x_{s,\cdot}$ are indistinguishable from classical derivatives $\partial^\alpha \hat{X}^x_{s,\cdot}$. In order to use the strong completeness Corollary 3.10 in [44], we show that for each $R > 0$, $k_1 > 2(n+1)$, it holds that

$$
\sup_{0\leq|\alpha|\leq p-1} \sup_{x,x'\in B_R} \sup_{s,s'\in[0,T]} \frac{\mathbb{E}\sup_{t\in[0,T]}|\partial^\alpha \hat{X}^{x'}_{s',t} - \partial^\alpha \hat{X}^x_{s,t}|^{k_1}}{(|x'-x|^2 + |s'-s|^2)^{\frac{\nu_1}{2}}} < \infty, \tag{4.6.3}
$$

where ν_1 is the same constant from Lemma 4.6.1. The marginal differences in x and s in the numerator are considered separately. By Lemma 4.3.2 or Theorem 4.3.3, the difference term in x in the numerator of (4.6.3) has the bound

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|\partial^\alpha \hat{X}^{x'}_{s',t} - \partial^\alpha \hat{X}^{x}_{s',t}|^{k_1}\Big] \leq C|x'-x|^{k_1}$$

for all $s \in [0,T]$, $x, x' \in B_R$, which is the desired Hölder bound for (4.6.3). For the difference term in s in the numerator of (4.6.3), it holds that

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,T]}|\partial^\alpha \hat{X}^{x}_{s',t} - \partial^\alpha \hat{X}^{x}_{s,t}|^{k_1}\Big] &\leq \mathbb{E}\Big[\sup_{t\in[s\wedge s',s\vee s']}|\partial^\alpha \hat{X}^{x}_{s\wedge s',t} - \partial^\alpha \hat{X}^{x}_{s\vee s',s\vee s'}|^{k_1}\Big] \\ &\quad + \mathbb{E}\Big[\sup_{t\in[s\vee s',T]}|\partial^\alpha \hat{X}^{x}_{s',t} - \partial^\alpha \hat{X}^{x}_{s,t}|^{k_1}\Big],\end{aligned} \tag{4.6.4}$$

where the first term on the right-hand side has the desired Hölder bound for (4.6.3) by Lemma 4.6.1. For the second term, by Assumption 10(iv) and Lemma 4.4.4, combined with Theorem 5.3 in [114], the joint system solved by $(\partial^\alpha X^x_{s,t})_{0\leq|\alpha|\leq p-1}$ is regular [114, Definition 2.1] and the same holds for the sum $(\partial^\alpha X^x_{s',t} - \partial^\alpha X^x_{s,t})_{0\leq|\alpha|\leq p-1}$ by an easy argument; therefore the strong Markov property (Theorem 2.13 in [114] with Proposition 4.1.5 in [63])[3] yields for any $R' > 0$ that

$$\begin{aligned}&\mathbb{E}\Big[\sup_{t\in[s\vee s',T]}|\partial^\alpha \hat{X}^{x}_{s',t} - \partial^\alpha \hat{X}^{x}_{s,t}|^{k_1}\wedge R'\Big] \\ &\quad = \mathbb{E}\Big[\Big[\sup_{t\in[s\vee s',T]}|\partial^\alpha \hat{X}^{x}_{s',t} - \partial^\alpha \hat{X}^{x}_{s,t}|^{k_1}\wedge R'\Big|\mathcal{F}_{s\vee s'}\Big]\Big] \\ &\quad = \int\!\!\int \sup_{t\in[s\vee s',T]}\Big|\partial^\alpha \hat{X}^{(\partial^\beta \hat{X}^x_{s\wedge s',s\vee s'}(\omega))_\beta}_{s\vee s',t}(\omega') - \partial^\alpha \hat{X}^{x}_{s\vee s',t}(\omega')\Big|^{k_1}\wedge R'\, d\mathbb{P}(\omega')d\mathbb{P}(\omega),\end{aligned} \tag{4.6.5}$$

where $\partial^\alpha \hat{X}^{(\partial^\beta \hat{X}^x_{s\wedge s',s\vee s'}(\omega))_\beta}_{s\vee s',t}(\omega')$ denotes the solution to the same (joint) system as $\partial^\alpha \hat{X}^x_{s\vee s',t}(\omega')$ but with initial conditions $\partial^\beta \hat{X}^x_{s\wedge s',s\vee s'}(\omega)$ for $0 \leq |\beta| \leq p-1$ for each respective partial derivative in place of the initial conditions x, e_i or 0. Then the proofs of Lemma 4.3.2 and Theorem 4.3.3 may be slightly modified in order to obtain analogous statements for the expectation in ω' in (4.6.5); the modification is namely that the initial condition (fixed with respect to ω') as mentioned can be added with no complications when Corollary 2.5

[3] Alternatively, we have uniqueness in the joint system by Theorem 3.5 in [111] and Theorem 1.2 in [114], so that Theorem IX.1.7 in [167], Itô's rule, Theorem 4.4.2 and Proposition 4.1.5 both in [63] give together the same required Markov property.

in [101] is applied. Given this, it holds that

$$\mathbb{E}\Big[\sup_{t\in[s\vee s',T]}|\partial^\alpha\hat{X}^x_{s',t}-\partial^\alpha\hat{X}^x_{s,t}|^{k_1}\wedge R'\Big]\leq C\sum_{\beta=0}^{|\alpha|-1}\mathbb{E}|\partial^\beta\hat{X}^x_{s\wedge s',s\vee s'}-\partial^\beta\hat{X}^x_{s\vee s',s\vee s'}|^{k_1},$$
$$=C\sum_{\beta=0}^{|\alpha|-1}\mathbb{E}|\partial^\beta\hat{X}^x_{s\wedge s',s\vee s'}-\partial^\beta\hat{X}^x_{s\wedge s',s\wedge s'}|^{k_1},$$

for all $x\in B_R$, $s,s'\in[0,T]$, $0\leq|\alpha|\leq p-1$, which, by Lemma 4.6.1 and dominated convergence in R', implies that the last term on the right-hand side of (4.6.4) has the desired Hölder bound for (4.6.3). Gathering the above and using the triangle inequality, (4.6.3) holds. Consequently, using on the way Lemma 4.3.2 and Theorem 4.3.3, Corollary 3.10 in [44] may be applied with $\beta=\frac{\nu_1}{k_1}$, $D=[0,T]\times\mathbb{R}^n$, $E=F=C([0,T],\mathbb{R}^n)$, $X=(\Omega\times[0,T]\times\mathbb{R}^n\ni(\omega,s,x)\mapsto\partial^\alpha\hat{X}^x_{s,\cdot}(\omega)\in C([0,T],\mathbb{R}^n))$ to obtain for $0\leq|\alpha|\leq p-1$ existence of an $\mathcal{F}\otimes\mathcal{B}(\mathbb{R}^n)$-measurable $\Omega\times\times[0,T]\times\mathbb{R}^n\ni(\omega,s,x)\mapsto\overline{\partial^\alpha X}^x_{s,\cdot}(\omega)\in C([0,T],\mathbb{R}^n)$ such that for all $\omega\in\Omega$, the function $[0,T]\times\mathbb{R}^n\ni(s,x)\mapsto\overline{\partial^\alpha X}^x_{s,\cdot}\in C([0,T],\mathbb{R}^n)$ is continuous and for any $(s,x)\in[0,T]\times\mathbb{R}^n$, $\overline{\partial^\alpha X}^x_{s,\cdot}$ is indistinguishable from $\partial^\alpha\hat{X}^x_{s,\cdot}$.
Since partial integrals of (jointly) continuous functions are still continuous, we may partially integrate $|\alpha|$ times each $\Delta_T\times\mathbb{R}^n\ni((s,t),x)\mapsto\overline{\partial^\alpha X}^x_{s,t}\in\mathbb{R}^n$ from 0 to x_i in order to obtain for each α,ω a continuous function $\Delta_T\times\mathbb{R}^n\ni((s,t),x)\mapsto\bar{X}^{x,\alpha}_{s,t}\in\mathbb{R}^n$, where along the way the continuous functions of the form $((s,t),x)\mapsto\overline{\partial^\beta X}^{(x_1,\dots,0,\dots,x_n)}_{s,t}$ are to be added in line with the fundamental theorem of calculus. For any $(s,t)\in\Delta_T$, by definition of $\overline{\partial^\alpha X}^x_{s,t}$ and its continuity in x, it holds $\mathbb{P}$-a.s. that $\overline{\partial^\alpha X}^x_{s,t}=\partial^\alpha\hat{X}^x_{s,t}$ for all $x\in\mathbb{R}^n$, so that their partial integrals in x are also $\mathbb{P}$-a.s. equal for all $x\in\mathbb{R}^n$ and in particular it holds $\mathbb{P}$-a.s. that $\bar{X}^{x,\alpha}_{s,t}=\hat{X}^{x,\alpha}_{s,t}$, for all $x\in\mathbb{R}^n$. Therefore, by continuity in $(s,t),x$, these functions coincide $\mathbb{P}$-a.s. across α, that is, it holds $\mathbb{P}$-a.s. that $\bar{X}^{x,\alpha}_{s,t}=\bar{X}^{x,\alpha'}_{s,t}$ and thus $\partial^\beta\bar{X}^{x,\alpha}_{s,t}=\partial^\beta\bar{X}^{x,\alpha'}_{s,t}$ for all $(s,t)\in\Delta_T$, $x\in\mathbb{R}^n$ and multi-indices α,α',β with $|\alpha|,|\alpha'|,|\beta|\in[0,p-1]$. Let this $\mathbb{P}$-a.s. defined function be denoted by $\bar{X}^x_{s,t}$, then the assertions about $\bar{X}^x_{s,t}$ in the statement of the lemma have been shown beside indistinguishability with the corresponding derivatives in probability, which holds by continuity in t for both functions.
For the last assertion, the Markov property as used earlier (this time only Theroem 2.13 in [114]) will be applied repeatedly without further mention. Since Assumption 10(ii) implies in particular for any $p^\dagger>0$ that

$$|b(x)|^{p^\dagger}+\|\sigma(x)\|^{p^\dagger}\leq Ce^{U(x)e^{-\rho t}}$$

for all $x \in \mathbb{R}^n$, $t \in [0,T]$, by Corollary 3.3 in [101] and Assumption 10(ii), it holds that

$$\begin{aligned}
&\sup_{\theta\in\Theta}\sup_{0\le s\le t\le T}\mathbb{E}[|b(\bar{X}^{Y^\theta_s}_{s,t})|^{p^\dagger}+\|\sigma(\bar{X}^{Y^\theta_s}_{s,t})\|^{p^\dagger}]\\
&\quad\le C\sup_{\theta\in\Theta}\sup_{0\le s\le t\le T}\mathbb{E}[e^{U(\bar{X}^{Y^\theta_s}_{s,t})e^{-\rho(t-s)}}]\\
&\quad\le C\sup_{\theta\in\Theta}\sup_{0\le s\le t\le T}\mathbb{E}[e^{U(\bar{X}^{Y^\theta_s}_{s,t})e^{-\rho(t-s)}+\int_s^t\bar{U}(\bar{X}^{Y_s}_{s,u})e^{-\rho(u-s)}du}]\\
&\quad\le C\sup_{\theta\in\Theta}\sup_{0\le s\le T}\mathbb{E}[e^{U(Y^\theta_s)}],
\end{aligned}$$

which is finite by Theorem 2.9 in [105]. For any $p^\dagger > 0$, by Assumption 10(ii), Corollary 3.3 in [101] and that $e^{-\rho(s-r)}, e^{-\rho r} < 1$, it holds that

$$\begin{aligned}
&\sup_{0\le r\le s\le t\le T}\mathbb{E}\left|\bar{X}^{\bar{X}^{Y^\theta_r}_{r,s}}_{t,T}\right|^{p^\dagger}\\
&\le C\sup_{0\le r\le s\le t\le T}\mathbb{E}[e^{U(\bar{X}^{\bar{X}^{Y^\theta_r}_{r,s}}_{t,T})e^{-\rho(T-t)}e^{-\rho(s-r)}e^{-\rho r}+\int_t^T\bar{U}(\bar{X}^{\bar{X}^{Y^\theta_r}_{r,s}}_{t,u})e^{-\rho(u-t)}e^{-\rho(s-r)}e^{-\rho r}du}]\\
&\le C\sup_{0\le r\le s\le T}\mathbb{E}[e^{U(\bar{X}^{Y^\theta_r}_{r,s})e^{-\rho(s-r)}e^{-\rho r}}]\\
&\le C\sup_{0\le r\le s\le T}\mathbb{E}[e^{U(\bar{X}^{Y^\theta_r}_{r,s})e^{-\rho(s-r)}e^{-\rho r}+\int_r^s\bar{U}(\bar{X}^{Y^\theta_r}_{r,u})e^{-\rho(u-r)}e^{-\rho r}du}]\\
&\le C\sup_{0\le r\le T}\mathbb{E}[e^{U(Y^\theta_r)e^{-\rho r}}],
\end{aligned}$$

for all $\theta \in \Theta$, which is finite uniformly in θ by Theorem 2.9 in [105].
For the higher derivatives, first note that for V_0 satisfying (4.6.1) and $0 < l < 1$, (4.6.1) is also satisfied with V_0^l in place of V_0. Moreover, the respective Lyapunov functions they generate satisfy Assumptions 7 and 8. Therefore, for any $\tilde{I} \in \mathbb{N} \cap [1, p-1]$, $\kappa \in \{(\kappa_i)_{i=1,\dots,\tilde{I}} : \kappa_i \in \mathbb{R}^n, |\kappa_i| = 1\}$, we may choose $l = \frac{2l^*}{\mathrm{degree}(q_0)}$, with q_0 from Theorem 4.3.3, so that for $\tilde{p}^\dagger > 0$, by Lemma 4.3.2 or Theorem 4.3.3, Young's inequality,

Assumptions 10(ii)(iv) and Theorem 2.4 in [101], it holds that

$$\begin{aligned}
&\sup_{0\le r\le s\le t\le T}\mathbb{E}\Big|\partial^{(\kappa)}X_{t,T}^{X_{r,s}^{Y_r^\theta}}\Big|^{\tilde{p}^\dagger}\\
&\le C\sup_{0\le r\le s\le T}\mathbb{E}[1+V(0,\bar{x}(X_{r,s}^{Y_r^\theta}))^{2l^*}+\sum_{i=1}^{i^*}\hat{V}_{l_i}^{0,T}(0,\hat{x}_{l_i}(X_{r,s}^{Y_r^\theta}))^{2l^*}]\\
&\le C\sup_{0\le r\le s\le T}\mathbb{E}[1+V(s-r,\bar{x}(Y_r^\theta))^{l^*}+\sum_{i=1}^{i^*}\hat{V}_{l_i}^{0,T}(s-r,\hat{x}_{l_i}(Y_r^\theta))^{l^*}]\\
&\le C\sup_{0\le r\le T}\mathbb{E}[1+V(0,\bar{x}(Y_r^\theta))^{l^*}+\sum_{i=1}^{i^*}\hat{V}_{l_i}^{0,T}(0,\hat{x}_{l_i}(Y_r^\theta))^{l^*}]\\
&\le C\sup_{0\le r\le T}\mathbb{E}[1+e^{U(Y_r^\theta)e^{-\rho T}}],
\end{aligned}$$

where C is in particular independent of $\kappa\in\{(\kappa_i)_{i=1,\dots,\tilde{I}}:\kappa_i\in\mathbb{R}^n,|\kappa_i|=1\}$ and $\theta\in\Theta$, so that the right-hand side is finite uniformly in θ by Theorem 2.9 in [105] and also uniformly in $\tilde{I}$.

□

The main theorem of this section about weak convergence of order 1 for the stopped increment-tamed Euler-Maruyama scheme is as follows.

Theorem 4.6.3. *Let Assumption 10 hold. For $f\in C^3(\mathbb{R}^n,\mathbb{R})$, if there exist constants $q^\dagger, C_f>0$ such that*

$$|\partial^\alpha f(x)|\le C_f(1+|x|^{q^\dagger})\tag{4.6.6}$$

for all $x\in\mathbb{R}^n$ and multiindices α with $0\le|\alpha|\le 3$, then there exists a constant $C>0$ such that

$$\Big|\mathbb{E}f(X_{0,T}^{Y_0^\theta})-\mathbb{E}f(Y_T^\theta)\Big|\le C\sup_{k\in\mathbb{N}_0\cap[0,n^*)}(t_{k+1}-t_k)$$

for all $\theta\in\Theta$, where $\theta=(t_0,\dots,t_{n^*})$.

Proof. Throughout the proof, we write $D_{|\theta|}=\{y:|y|<\exp(|\log\sup_k t_{k+1}-t_k|^{\frac{1}{2}})\}$. To begin, we rewrite the approximation Y_t^θ as an SDE. For every $k\in\mathbb{N}_0\cap[0,n^*-1]$, $\theta=(t_0,\dots,t_{n^*})\in\Theta$, consider

$$\begin{aligned}
Z_t^{\theta,k}&=\begin{cases}0 & \text{if } t<t_k\\ b(Y_{t_k}^\theta)(t-t_k)+\sigma(Y_{t_k}^\theta)(W_t-W_{t_k}) & \text{if } t_k\le t<t_{k+1}\\ b(Y_{t_k}^\theta)(t_{k+1}-t_k)+\sigma(Y_{t_k}^\theta)(W_{t_{k+1}}-W_{t_k}) & \text{if } t_{k+1}\le t\end{cases}\\
&=\int_0^t \mathbb{1}_{[t_k,t_{k+1})}(u)b(Y_{t_k}^\theta)du+\int_0^t \mathbb{1}_{[t_k,t_{k+1})}(u)\sigma(Y_{t_k}^\theta)dW_u,
\end{aligned}\tag{4.6.7}$$

defined for all $t \in [0,T]$, then Y_t^θ solves

$$Y_t^\theta = Y_0^\theta + \sum_{k=0}^{n^*-1} \mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta) \frac{Z_t^{\theta,k}}{1+|Z_t^{\theta,k}|^{q'}}, \tag{4.6.8}$$

where by Itô's rule, for $\hat{f} : \mathbb{R}^n \to \mathbb{R}^n$ given by $\hat{f}(z) = \frac{z}{1+|z|^{q'}}$, it holds that

$$\begin{aligned}\frac{Z_t^{\theta,k}}{1+|Z_t^{\theta,k}|^{q'}} &= \int_0^t \mathbb{1}_{[t_k,t_{k+1})}(u)(b(Y_{t_k}^\theta) + b^*(Y_{t_k}^\theta, Z_u^{\theta,k}))du \\ &\quad + \int_0^t \mathbb{1}_{[t_k,t_{k+1})}(u)(\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_u^{\theta,k}))dW_u \end{aligned}\tag{4.6.9}$$

and $b^* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma^* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ are given by

$$b^*(y,z) = -b(y)\bigg(\frac{|z|^{q'}}{1+|z|^{q'}}\bigg) - q'z\bigg(z\cdot b(y)\frac{|z|^{q'-2}}{(1+|z|^{q'})^2}\bigg) + \frac{1}{2}((\sigma\sigma^\top(y)) : D^2)\hat{f}(z) \tag{4.6.10}$$

$$\sigma^*(y,z) = -\sigma(y)\bigg(\frac{|z|^{q'}}{1+|z|^{q'}}\bigg) - q'z\bigg(z^\top\sigma(y)\frac{|z|^{q'-2}}{(1+|z|^{q'})^2}\bigg). \tag{4.6.11}$$

Note that using $q' \geq 3$, there exists a constant $\nu_2 \geq 2$ such that the second order derivatives satisfy $|\partial_{ij}^2 \hat{f}(z)| \leq C|z|^{\nu_2}$ for all $z \in \mathbb{R}^n$, $i,j \in \mathbb{N}\cap[1,n]$.
By Theorem 3.1 in [99] and Lemma 4.6.2, for any $\theta \in \Theta$, it holds that

$$\begin{aligned}&\mathbb{E}[f(X_{0,T}^{Y_0^\theta})] - \mathbb{E}[f(Y_T^\theta)] \\ &= \sum_{k=0}^{n^*-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\bigg(\bigg(\bigg(b(Y_t^\theta) - \mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)(b(Y_{t_k}^\theta) \\ &\quad + b^*(Y_{t_k}^\theta, Z_t^{\theta,k}))\bigg)\cdot\nabla\bigg)\bar{X}_{t,T}^{Y_t^\theta}\bigg)\cdot\nabla\bigg)f(\bar{X}_{t,T}^{Y_t^\theta})dt \\ &\quad + \frac{1}{2}\mathbb{E}\int_{t_k}^{t_{k+1}}\sum_{i,j=1}^n\Big(\sigma(Y_t^\theta)\sigma(Y_t^\theta)^\top - \mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)(\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k}))(\sigma(Y_{t_k}^\theta) \\ &\quad + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k}))^\top\Big)_{ij}\Big(((\partial_i\bar{X}_{t,T}^{Y_t^\theta}\otimes\partial_j\bar{X}_{t,T}^{Y_t^\theta}) : D^2)f(\bar{X}_{t,T}^{Y_t^\theta}) + (\partial_{ij}^2\bar{X}_{t,T}^{Y_t^\theta}\cdot\nabla)f(\bar{X}_{t,T}^{Y_t^\theta})\Big)dt\end{aligned} \tag{4.6.12}$$

For the first terms on the right-hand side of (4.6.12), denoting

$$\hat{b}^*(y',y,z) = b(y') - \mathbb{1}_{D_{|\theta|}}(y)(b(y) + b^*(y,z)), \tag{4.6.13}$$

it holds that

$$
\begin{aligned}
&\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k})\cdot\nabla\Big)\bar{X}_{t,T}^{Y_t^\theta}\Big)\cdot\nabla\Big)f(\bar{X}_{t,T}^{Y_t^\theta})\\
&\quad=\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k})\cdot\nabla\Big)\Big(\bar{X}_{t,T}^{Y_t^\theta}-\bar{X}_{t,T}^{Y_{t_k}^\theta}\Big)\Big)\cdot\nabla\Big)f(\bar{X}_{t,T}^{Y_t^\theta})\\
&\qquad+\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k})\cdot\nabla\Big)\bar{X}_{t,T}^{Y_{t_k}^\theta}\Big)\cdot\nabla\Big)\Big(f(\bar{X}_{t,T}^{Y_t^\theta})-f(\bar{X}_{t,T}^{Y_{t_k}^\theta})\Big)\\
&\qquad+\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k})\cdot\nabla\Big)\bar{X}_{t,T}^{Y_{t_k}^\theta}\Big)\cdot\nabla\Big)f(\bar{X}_{t,T}^{Y_{t_k}^\theta}).
\end{aligned}
\tag{4.6.14}
$$

The first part of the factor involving b has the form

$$
\begin{aligned}
&b(Y_t^\theta)-\mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)b(Y_{t_k}^\theta)\\
&\quad=\Big[b(Y_t^\theta)-b(Y_{t_k}^\theta)\Big]+\Big[b(Y_{t_k}^\theta)-\mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)b(Y_{t_k}^\theta)\Big]\\
&\quad=\int_{t_k}^t \mathbb{1}_{D_\theta}(Y_{t_k})\Big(\Big((b(Y_{t_k}^\theta)+b^*(Y_{t_k}^\theta, Z_u^{\theta,k}))\cdot\nabla\Big)b(Y_u^\theta)\\
&\qquad+\frac{1}{2}\Big(\Big((\sigma(Y_{t_k}^\theta)+\sigma^*(Y_{t_k}^\theta, Z_u^{\theta,k}))(\sigma(Y_{t_k}^\theta)+\sigma^*(Y_{t_k}^\theta, Z_u^{\theta,k}))^\top\Big):D^2\Big)b(Y_u^\theta)\Big)du\\
&\qquad+\int_{t_k}^t \mathbb{1}_{D_\theta}(Y_{t_k})\Big((\sigma(Y_{t_k}^\theta)+\sigma^*(Y_{t_k}^\theta, Z_u^{\theta,k}))\cdot\nabla\Big)b(Y_u^\theta)dW_u\\
&\qquad+b(Y_{t_k}^\theta)(1-\mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)),
\end{aligned}
\tag{4.6.15}
$$

where the integral w.r.t. u is uniformly bounded in θ by $C(t-t_k)$ in $L^2(\mathbb{P})$ norm, the stochastic integral is uniformly bounded in θ by $C(t-t_k)^{\frac{1}{2}}$ in $L^2(\mathbb{P})$ norm and the last term has the same property as the integral-in-u (and in fact of arbitrary order in $t-t_k$) by the calculation of inequalities (47), (48) in [103]. Using the definition (4.6.10) for b^* along with $q'\geq 3$, there exists a constant $\nu_2\geq 2$ such that the remaining part of the factor involving b from (4.6.14) has the bound

$$
|\mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)b^*(Y_{t_k}^\theta, Z_t^{\theta,k})|\leq C|b(Y_{t_k}^\theta)||Z_t^{\theta,k}|^{\nu_2}
\tag{4.6.16}
$$

for all $\theta\in\Theta$. Putting (4.6.15) and (4.6.16) into the first term on the right-hand side of (4.6.14) and using Hölder's inequality, Assumptions 10(ii)(iv), together with the equalities (4.6.7), (4.6.8), (4.6.9), (4.6.10), (4.6.11), inequality (4.6.6), Lemma 4.6.2, Lemma 4.3.2, Theorem 4.3.3, Markov property (Theorem 2.13 in [114]; see also justification in the proof of Lemma 4.6.2) and exponential integrability for U as in Theorem 2.9 in [105] yield

$$
\mathbb{E}\Big|\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k})\cdot\nabla\Big)\Big(\bar{X}_{t,T}^{Y_t^\theta}-\bar{X}_{t,T}^{Y_{t_k}^\theta}\Big)\Big)\cdot\nabla\Big)f(\bar{X}_{t,T}^{Y_t^\theta})\Big|\leq C(t-t_k)
\tag{4.6.17}
$$

for all $t \in [t_k, t_{k+1})$, $\theta \in \Theta$. The same arguments can be used for the second term on the right-hand side of (4.6.14), along with the additional estimate

$$
\begin{aligned}
&\mathbb{E}|\partial_i f(\bar{X}_{t,T}^{Y_t^\theta}) - \partial_i f(\bar{X}_{t,T}^{Y_{t_k}^\theta})|^r \\
&\quad \le \mathbb{E}\bigg| \int_0^1 \nabla \partial_i f(\lambda \bar{X}_{t,T}^{Y_t^\theta} + (1-\lambda)\bar{X}_{t,T}^{Y_{t_k}^\theta}) d\lambda \cdot (\bar{X}_{t,T}^{Y_t^\theta} - \bar{X}_{t,T}^{Y_{t_k}^\theta}) \bigg|^r \\
&\quad \le C(1 + \mathbb{E}|\bar{X}_{t,T}^{Y_t^\theta}|^{2q^\dagger} + \mathbb{E}|\bar{X}_{t,T}^{Y_{t_k}^\theta}|^{2q^\dagger})^{\frac{r}{2}} (\mathbb{E}|\bar{X}_{t,T}^{Y_t^\theta} - \bar{X}_{t,T}^{Y_{t_k}^\theta}|^2)^{\frac{r}{2}} \\
&\quad \le C(\mathbb{E}e^{U(Y_t^\theta)} + \mathbb{E}e^{U(Y_{t_k}^\theta)}) \mathbb{E}|Y_t^\theta - Y_{t_k}^\theta|^r \\
&\quad \le C(t - t_k)^{\frac{r}{2}}
\end{aligned}
$$

where $r > 1$, in order to obtain the same right-hand bound as (4.6.17). For the last term on the right-hand side of (4.6.14), we rely more prominently on the Markov property. For any $R > 0$, it holds that

$$
\begin{aligned}
&\mathbb{E}\bigg[\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k}) \cdot \nabla \Big) \bar{X}_{t,T}^{Y_{t_k}^\theta} \Big) \cdot \nabla \Big) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}) \wedge R \bigg] \\
&\quad = \mathbb{E}\bigg[\mathbb{E}\bigg[\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k}) \cdot \nabla \Big) \bar{X}_{t,T}^{Y_{t_k}^\theta} \Big) \cdot \nabla \Big) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}) \wedge R \bigg| \mathcal{F}_t \bigg] \bigg] \\
&\quad = \sum_{i=1}^n \mathbb{E}\bigg[\hat{b}_i^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k}) \mathbb{E}\Big[\Big(\partial_i \bar{X}_{t,T}^{Y_{t_k}^\theta} \cdot \nabla \Big) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}) \wedge R \Big| \mathcal{F}_t \Big] \bigg] \\
&\quad = \sum_{i=1}^n \mathbb{E}\bigg[\mathbb{E}\Big[\hat{b}_i^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k}) \Big| \mathcal{F}_{t_k} \Big] \mathbb{E}\Big[\Big(\partial_i \bar{X}_{t,T}^{Y_{t_k}^\theta} \cdot \nabla \Big) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}) \wedge R \Big| \mathcal{F}_{t_k} \Big] \bigg],
\end{aligned}
$$

so that (4.6.13), (4.6.15) and (4.6.16), where the only order $\frac{1}{2}$ term in $t - t_k$ from (4.6.15) has vanished, together with the same arguments as before and dominated convergence in R yields

$$
\mathbb{E}\bigg[\Big(\Big(\Big(\hat{b}^*(Y_t^\theta, Y_{t_k}^\theta, Z_t^{\theta,k}) \cdot \nabla \Big) \bar{X}_{t,T}^{Y_{t_k}^\theta} \Big) \cdot \nabla \Big) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}) \bigg] \le C(t - t_k) \tag{4.6.18}
$$

for all $t \in [t_k, t_{k+1})$, $\theta \in \Theta$. Gathering the arguments from (4.6.17) onwards, the integrals involving b in (4.6.12) have been shown to be of order $t - t_k$. For the integrals involving σ in (4.6.12), after rewriting

$$
\begin{aligned}
&\sigma(Y_t^\theta)\sigma(Y_t^\theta)^\top - \mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)(\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k}))(\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k}))^\top \\
&\quad = \Big(\sigma(Y_t^\theta) - \mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)(\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k})) \Big) \sigma(Y_t^\theta)^\top \\
&\qquad + \mathbb{1}_{D_{|\theta|}}(Y_{t_k}^\theta)(\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k})) \Big(\sigma(Y_t^\theta)^\top - (\sigma(Y_{t_k}^\theta) + \sigma^*(Y_{t_k}^\theta, Z_t^{\theta,k}))^\top \Big)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&((\partial_i \bar{X}_{t,T}^{Y_t^\theta} \otimes \partial_j \bar{X}_{t,T}^{Y_t^\theta}) : D^2) f(\bar{X}_{t,T}^{Y_t^\theta}) + (\partial_{ij}^2 \bar{X}_{t,T}^{Y_t^\theta} \cdot \nabla) f(\bar{X}_{t,T}^{Y_t^\theta}) \\
&\quad = (((\partial_i \bar{X}_{t,T}^{Y_t^\theta} - \partial_i \bar{X}_{t,T}^{Y_{t_k}^\theta}) \otimes \partial_j \bar{X}_{t,T}^{Y_t^\theta}) : D^2) f(\bar{X}_{t,T}^{Y_t^\theta}) + ((\partial_{ij}^2 \bar{X}_{t,T}^{Y_t^\theta} - \partial_{ij}^2 \bar{X}_{t,T}^{Y_{t_k}^\theta}) \cdot \nabla) f(\bar{X}_{t,T}^{Y_t^\theta}) \\
&\qquad + (((\partial_i \bar{X}_{t,T}^{Y_t^\theta} - \partial_i \bar{X}_{t,T}^{Y_{t_k}^\theta}) \otimes \partial_j \bar{X}_{t,T}^{Y_{t_k}^\theta}) : D^2) f(\bar{X}_{t,T}^{Y_t^\theta}) + (\partial_{ij}^2 \bar{X}_{t,T}^{Y_{t_k}^\theta} \cdot \nabla)(f(\bar{X}_{t,T}^{Y_t^\theta}) - f(\bar{X}_{t,T}^{Y_{t_k}^\theta})) \\
&\qquad + ((\partial_i \bar{X}_{t,T}^{Y_{t_k}^\theta} \otimes \partial_j \bar{X}_{t,T}^{Y_{t_k}^\theta}) : D^2)(f(\bar{X}_{t,T}^{Y_t^\theta}) - f(\bar{X}_{t,T}^{Y_{t_k}^\theta})) + (\partial_{ij}^2 \bar{X}_{t,T}^{Y_{t_k}^\theta} \cdot \nabla) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}) \\
&\qquad + ((\partial_i \bar{X}_{t,T}^{Y_{t_k}^\theta} \otimes \partial_j \bar{X}_{t,T}^{Y_{t_k}^\theta}) : D^2) f(\bar{X}_{t,T}^{Y_{t_k}^\theta}),
\end{aligned}
$$

the same bound as (4.6.18) holds for all of (4.6.12) by the same treatment as for (4.6.18). □

4.7 Examples

In this section, specific examples are provided where the results presented above are applicable. As stated in the introduction, most of the examples in [44, 105] are viable and many Lyapunov functions have already been given in these references (applicable here after a simple transformation, see Remark 4.6.1). Here, the focus is placed on two particular examples differing in some considerable way to the aforementioned references. In Section 4.7.1, our results are applied to the (underdamped) Langevin dynamics with variable friction, which, by definition, does not have globally Lipschitz (nor monotone) coefficients.In Section 4.7.2, a Lyapunov function of the classical type (that is, V_0 satisfying $LV_0 \leq CV_0$) is given for the Stochastic Duffing-van der Pol equation; moreover, this is given in consideration of a limiting parameter case that has not fallen under the assumptions in previous works mentioned above.

4.7.1 Langevin equation with variable friction

Here, the backward Kolmogorov equation and Poisson equation associated with the Langevin equation are shown to hold even in cases where the friction matrix depends on both position and velocity variables. The pointwise solution to the Kolmogorov equation is used to obtain a distributional solution to the associated Poisson equation and in doing so the derivation of a gradient formula for the asymptotic variance as in the previous chapter is made viable; the last part is not explored further here. Note that the case where the friction matrix depends on the velocity variable was considered in [11]. We allow the potential to not be infinitely differentiable and do not make use of hypoellipticity.

Assumption 11. The function $U \in C^3(\mathbb{R}^n)$ is such that there exists $\tilde{k}, \tilde{K} > 0$ with $\nabla U(q) \cdot q \geq \tilde{k}|x|^2 - \tilde{K}$ for all $q \in \mathbb{R}^n$. The friction matrix $\Gamma \in C^\infty(\mathbb{R}^{2n}, \mathbb{R}^{n\times n}) \cap L^\infty$

is symmetric positive definite everywhere such that there exist[4] $\beta_1 < 1$, $\tilde{m}, \tilde{M} > 0$ with $|\nabla_p \cdot \Gamma(q,p)| < \tilde{M}(1 + |q|^{\beta_1} + |p|^{\beta_1})$ and $\Gamma(q,p) \geq \tilde{m}I$ for all $q,p \in \mathbb{R}^n$.

Note Assumption 11 implies for $R > 1$, $q \in \mathbb{R}^n$ with $|q| = 1$,

$$\begin{aligned} U(Rq) - U(q) &= \int_1^R \nabla U(\lambda q) \cdot \frac{\lambda q}{\lambda} d\lambda \geq \int_1^R (\tilde{k}|\lambda q|^2 - \tilde{K})\lambda^{-1} d\lambda \\ &= \frac{\tilde{k}(R^2 - 1)}{2} - \tilde{K} \log R, \end{aligned}$$

which yields $U(q) \geq \frac{\tilde{k}}{4}|q|^2 - C$ for all $q \in \mathbb{R}^n$ and some constant $C > 0$. Consider $\mathbb{R}^{2n}$-valued solutions (q_t, p_t) to

$$dq_t = p_t dt \tag{4.7.1a}$$

$$dp_t = -\nabla U(q_t)dt + \nabla_p \cdot \Gamma(q_t,p_t)dt - \Gamma(q_t,p_t)p_t dt + \sqrt{\Gamma}(q_t,p_t)dW_t, \tag{4.7.1b}$$

where $\sqrt{\Gamma}$ denotes some matrix satisfying $\sqrt{\Gamma}\sqrt{\Gamma}^\top = \Gamma$ and $(\nabla \cdot \Gamma)_i = \sum_j \nabla_{p_j}\Gamma_{ij}$. For $b = \min(\tilde{k}^{-1}(\sup_{\mathbb{R}^{2n}}|\Gamma|)^{-1}, \tilde{m}, \tilde{k}^{\frac{1}{2}})$, $a = \frac{1}{4}\min(\frac{b}{\tilde{k}}, \tilde{m})$, let

$$V_\gamma(q,p) = e^{\gamma(U(q) + a|q|^2 + bq\cdot p + |p|^2)}. \tag{4.7.2}$$

In the following, $|M|$ denotes the operator norm of $M \in \mathbb{R}^{n\times n}$.

Proposition 4.7.1. *Under Assumption 11, there exists constants $c_1, c_2, c_3 > 0$ such that for all γ satisfying*

$$0 < \gamma \leq \gamma^* := \frac{1}{8}\min((\tilde{k}b\sup_{\mathbb{R}^{2n}}|\Gamma|)^{-1}, \tilde{m}(4\sup_{\mathbb{R}^{2n}}|\Gamma|)^{-1}), \tag{4.7.3}$$

it holds that

$$LV_\gamma(q,p) \leq (c_1 - c_2|q|^2 - c_3|p|^2)\gamma V_\gamma(q,p) \tag{4.7.4}$$

for all $(q,p) \in \mathbb{R}^{2n}$, where L is the generator (4.2.2) associated with (4.7.1).
If in addition there exist $0 < \beta_2 < 1$, $\bar{M} > 0$ such that

$$\begin{aligned} |\partial_i(\nabla_p \cdot \Gamma(q,p) - \nabla U(q))| &\leq \bar{M}(1 - \inf U + U(q) + |p|^2)^{\beta_2} \\ |\partial_i \Gamma(q,p)| &\leq \bar{M}(1 - \inf U^{\frac{1}{2}} + U(q)^{\frac{1}{2}} + |p|)^{\beta_2} \\ |\partial_i\partial_j(\nabla_p \cdot \Gamma(q,p) - \nabla U(q))| + |\partial_i\partial_j\Gamma(q,p)| &\leq \bar{M}(1 + e^{(U(q)+|p|^2)^{\beta_2}}) \end{aligned}$$

for all $q,p \in \mathbb{R}^n$, $i,j \in \{1,\dots,2n\}$, then Assumption 8 (with $p = 2$) is satisfied with

[4] It is possible to allow for $\beta_1 = 1$, but at the cost of more stringent bounds on the coefficients.

$\hat{V}_k = V_\gamma$ *with any* γ *satisfying* (4.7.3), $G(q,p) = C(1 - \inf U + U(q) + |p|^2)^{\beta_3}$ *for some constants* $C > 0$ *and* $\beta_2 < \beta_3 < 1$.

Proof. The left-hand side of (4.7.4) calculates as

$$
\begin{aligned}
&(p \cdot \nabla_q - \nabla_q U(q) \cdot \nabla_p + (\nabla_p \cdot \Gamma(q,p)) \cdot \nabla_p - (\Gamma(q,p)p) \cdot \nabla_p + \Gamma(q,p) : D^2) V_\gamma(q,p) \\
&\quad = (2aq \cdot p + b|p|^2 - b\nabla_q U(q) \cdot q + (\nabla_p \cdot \Gamma(q,p) - \Gamma(q,p)p) \cdot (bq + 2p) \\
&\qquad + 2\mathrm{Tr}\Gamma(q,p) + \gamma\Gamma(q,p) : (b^2 qq^\top + 4pp^\top))\gamma V_\gamma(q,p) \\
&\quad \leq \Bigg(\left(a - \frac{b}{\tilde{k}} + \frac{1}{2} b^2 |\Gamma| + b^2\gamma|\Gamma| \right) |q|^2 + \left(a + b + \frac{1}{2}|\Gamma| - 2\tilde{m} + 4\gamma|\Gamma| \right) |p|^2 \\
&\qquad + \tilde{M}(1 + |q|^{\beta_1} + |p|^{\beta_2})|bq + 2p| + b\tilde{K} + 2\mathrm{Tr}\Gamma \Bigg) \gamma V_\gamma(q,p) \\
&\quad \leq \left(c - \frac{b}{16\tilde{k}}|q|^2 - \frac{\tilde{m}}{16}|p|^2 \right) \gamma V_\gamma(q,p)
\end{aligned}
\tag{4.7.5}
$$

for some constant $c > 0$. The last assertion follows by straightforward applications of Young's inequality. $\square$

For U with locally Lipschitz third derivatives and by Theorem 4.5.2 (iii), the associated Poisson equation with right-hand side $\hat{f} = f - \int_{\mathbb{R}^{2n}} f d\mu \in L^2(\mu)$ holds in the distributional sense if in addition

$$
\left| \mathbb{E}\hat{f}(z_t^\cdot) \right| + \left| \int_t^\infty \mathbb{E}\hat{f}(z_s^\cdot) ds \right| \to 0 \text{ in } L^2(\mu) \text{ as } t \to \infty, \tag{4.7.6}
$$

where for any $z \in \mathbb{R}^{2n}$, $z_t^z = (q_t, p_t)$ solves (4.7.1), $\mathbb{P}((q_0,p_0) = z) = 1$ and $\mu(dq,dp) = Z^{-1}e^{-U(q) - \frac{p^2}{2}} dqdp$ is the invariant probability measure with normalizing constant Z. We obtain (4.7.6) in the following by using the ergodicity results of [52], see alternatively Theorem 2.4 in [194]. The proof of Proposition 1.2 in [194] can be modified for (4.7.1) to obtain

Proposition 4.7.2. *For every* $x \in \mathbb{R}^{2n}$, $t > 0$, *the measure* $P^t(x,\cdot)$ *admits a density* $p_t(x,\cdot)$ *satisfying* $p_t(x,y) > 0$ *Lebesgue almost every* $y \in \mathbb{R}^{2n}$ *and*

$$
(x \mapsto p_t(x,\cdot)) \in C(\mathbb{R}^{2n}, L^1(\mathbb{R}^{2n})). \tag{4.7.7}
$$

Proof. For the Markov property, see the proof of Lemma 4.6.2 just before (4.6.5). The proof in the aforementioned reference follows through except in the proof of Lemma 1.1 in [194], where the Lyapunov function (4.7.2) is to be used in place of $\tilde{H}(x,y) = \frac{1}{2}|y|^2 + V(x) - \inf_{\mathbb{R}^n} V + 1$ and R^2 in the ensuing calculations is replaced as needed. $\square$

Proposition 4.7.2 implies the existence of an irreducible skeleton chain, see [52]; together

with Theorem 3.2 in the same reference (with $\Psi = (\Psi_1, \Psi_2)$, $\Psi_1(x) = \Psi_2(x) = (x/2)^{\frac{1}{2}}$, $\phi(x) = x^{\frac{1}{2}}$, $V = V_{\frac{1}{2}\gamma^*}$), Theorem 3.4 in [52] (compact sets are petite by Theorem 4.1(i) in [137]) and Proposition 3.1 in the same reference (with $\phi(x) = x$, $V = V_{\gamma^*}$), this yields (4.7.6) for $\hat{f}$ satisfying $\hat{f}/V_{\frac{1}{8}\gamma^*} \in L^\infty$. Note that the Foster-Lyapunov condition for geometric ergodicity suffice as well.
In addition, maximal dissipativity of the closure of the generator defined on C_c^∞ is also enough to conclude a distributional solution to the Poisson equation, which motivates the question of whether there is a relationship between this property and the Kolmogorov equation; a partial answer is given by Proposition 4.4.6. However, maximal dissipativity is not generally available and for example not established for (4.7.1) with Γ dependent on both q and p, see [11] on the other hand.

4.7.2 Stochastic Duffing-van der Pol equation

We show here that the Stochastic Duffing-van der Pol oscillator admits a Lyapunov function satisfying the assumptions of Theorem 4.3.3. Note that in doing so, the difficult parts of Assumption 10 are shown to be satisfied, so that our Theorem 4.6.3 about weak numerical convergence rates applies. In particular, the logarithm of the Lyapunov function described below may be used for U in Assumption 10. The version of the equation considered is from [102] with $\beta_2 = 0$, which is less general than in [102] but still includes the setting of Section 13.1 in [112] and [5] for example. Moreover, it is more general than Section 4.3 in [44], which is reflected in the form of the Lyapunov function here. Note on the other hand, it is not more general than in [103]. Specifically, for $(W^{(1)}, W^{(3)}) : [0,T] \times \Omega \to \mathbb{R}^2$ a standard $(\mathcal{F}_t)_{t\in[0,T]}$-adapted Brownian motion, $\alpha_1, \alpha_2, \beta_1, \beta_3 \in \mathbb{R}$, $\alpha_3 \geq 0$, consider $\mathbb{R}^2$-valued solutions to (4.1.2), restated here for convenience:

$$dX_t^{(1)} = X_t^{(2)} dt, \tag{4.7.8a}$$
$$\begin{aligned} dX_t^{(2)} &= [\alpha_1 X_t^{(1)} - \alpha_2 X_t^{(2)} - \alpha_3 X_t^{(2)} (X_t^{(1)})^2 - (X_t^{(1)})^3] dt \\ &\quad + \beta_1 X_t^{(1)} dW_t^{(1)} + \beta_3 dW_t^{(3)}. \end{aligned} \tag{4.7.8b}$$

For $\alpha_3 > 0$ (see Remark 4.7.1 for the case $\alpha_3 = 0$, V is chosen here with regard to this case), let $a = \min(1, \frac{1}{\alpha_3})$, $b = (2-\alpha_3)\mathbb{1}_{\alpha_3<1} + \frac{3}{2}\mathbb{1}_{\alpha_3\geq 1}$, $c = 6|\alpha_2|$, $\gamma \leq \min(\frac{\alpha_3}{4\beta_1^2}, \frac{1}{\beta_1^2}, \frac{|\alpha_2|}{8\beta_3^2})$ and let $\eta : \mathbb{R} \to [0,1]$ be a C_c^∞ cut off function satisfying $\eta(y) = 1$ for $y^2 \leq \frac{1+|a-2\alpha_2 b+2\beta_3^2\gamma b^2|}{2\alpha_3 b-2\beta_1^2\gamma b^2}$. Define

$$\begin{aligned} V(x_1, x_2) &= V_1(x_1, x_2) + V_2(x_1, x_2) \\ &:= (1-\eta(x_1)) e^{\gamma(x_1^4 + a x_1 x_2 + b x_2^2)} + e^{\gamma(-c x_1 x_2 + \frac{1}{2} x_2^2)}. \end{aligned}$$

Proposition 4.7.3. *If $\alpha_3 > 0$, then there exists a constant $C^* > 0$ such that $LV \leq C^* V$, where L is the generator (4.2.2) associated with (4.7.8). Moreover, Assumptions 7 and 8 are satisfied with $G(t,x) = (3 + 2\sum_i |\alpha_i| + \beta_1^2)(1 + |x_1|^3 + |x_2|^{\frac{3}{2}})$ and $\hat{V}_k(t,x) = (|X_t^{(1)}|^4 + 2|X_t^{(2)}|^2 + 1)^k$ for $t \geq 0$, $x = (x_1, x_2) \in \mathbb{R}^n$, where $(X_t^{(1)}, X_t^{(2)})$ solves (4.7.8) with $(X_t^{(1)}, X_t^{(2)}) = (x_1, x_2)$.*

Proof. The functions V_1 and V_2 satisfy

$$
\begin{aligned}
LV_1(x_1,x_2) = {} & [(2\alpha_1 b - \alpha_2 a)x_1 x_2 + (a - 2\alpha_2 b + 2\beta_3^2\gamma b^2)x_2^2 + (\alpha_1 a + \frac{1}{2}\beta_3^2\gamma a^2 + \beta_1^2 b)x_1^2 \\
& + (2\beta_1^2\gamma b^2 - 2\alpha_3 b)x_1^2 x_2^2 - (\alpha_3 a + 2b - 4)x_1^3 x_2 + (\frac{1}{2}\beta_1^2\gamma a^2 - a)x_1^4 \\
& + b\beta_3^2 - \frac{x_2 \partial_{x_1}\eta(x_1)}{1-\eta(x_1)}]\gamma V_1(x_1,x_2) \\
LV_2(x_1,x_2) = {} & [(\frac{1}{2}\beta_3^2\gamma - c - \alpha_2)x_2^2 + (\frac{1}{2}c^2\beta_3^2\gamma - \alpha_1 c + \frac{1}{2}\beta_1^2)x_1^2 \\
& + (\alpha_2 c + \alpha_1)x_1 x_2 + (\alpha_3 c - 1)x_1^3 x_2 + (c + \frac{1}{2}c^2\gamma\beta_1^2)x_1^4 \\
& + (\frac{1}{2}\beta_1^2\gamma - \alpha_3)x_1^2 x_2^2 + \frac{1}{2}\beta_3^2]\gamma V_2(x_1,x_2).
\end{aligned}
$$

where $\frac{1}{1-\eta(x_1)} := 0$ whenever $1 - \eta(x_1) = 0$. In order to see $LV \leq CV$, consider separately the regimes $x_1^2 \leq \frac{1+|a-2\alpha_2 b+2\beta_3^2\gamma b^2|}{2\alpha_3 b - 2\beta_1^2\gamma b^2}$ and its complement in $\mathbb{R}^2$. In the former case, $V_1(x_1,x_2) = LV_1(x_1,x_2) = 0$ and by our choice of c and γ, there exists a generic constant $C > 0$ such that $LV_2 \leq CV_2$, therefore $LV \leq CV$. Otherwise in the complementary case where $|x_1|$ is bounded below, we have $LV_1 \leq CV_1$ and when in addition $x_1 \in \mathrm{supp}\eta \cup B_1(0)$, it holds that $LV_2 \leq CV_2$. It remains to estimate LV_2 when $x_1 \notin \mathrm{supp}\eta \cup B_1(0)$, in which case we have $|x|^i e^{\gamma(-cx_1x_2 + \frac{1}{2}x_2^2)} \leq Ce^{\gamma(\frac{1}{2}x_1^4 + \frac{3}{4}x_2^2)} \leq CV_1(x_1,x_2)$ for $i \leq 4$, from which $LV_2 \leq CV_1$.

For the second assertion, it is straightforward to see that (4.3.1), (4.3.2) hold and that the higher derivatives of the coefficients of (4.7.8) are bounded above in terms of $\hat{V}_k$ for any k, p as called-for in Assumption 8; for (4.3.3), consider separately the cases $|x_1| \leq \sup\{|x| : x \in \mathrm{supp}\eta\}$ and otherwise. In the former case, it holds that

$$
G(x_1,x_2) \leq C(1 + |x_2|^{\frac{3}{2}}),
$$

which yields that for any $m > 0$, there is $M = M(m) > 0$ continuous in m such that

$$
G \leq m\log(V_2) + M \leq m\log(V) + M. \tag{4.7.9}
$$

When $|x_1| > \sup\{|x| : x \in \mathrm{supp}\eta\}$, inequalities (4.7.9) continue to hold with V_1 replacing V_2 and a corresponding continuous function $m \mapsto M(m)$. $\square$

Remark 4.7.1. When $\alpha_3 = 0$, the arguments of Proposition 4.7.3 fail to adapt, but one can consider (4.7.8) and its derivative dynamics appended with $d\alpha_3 = 0$ with initial condition as a positive function in $\mathbb{R}$ in order to obtain the statements of Theorem 4.3.3. To elaborate, for almost all $\omega \in \Omega$, the solution to the appended derivative dynamics is continuously differentiable with respect to α_3 by Theorem V.39 in [165] and in particular is continuous, so that Fatou's lemma can be used to obtain

$$\mathbb{E} \lim_{\alpha_3 \to 0} \sup_{0 \leq s \leq t} \left| \partial^{(\kappa)} X_s^{x'} - \partial^{(\kappa)} X_s^x \right|^{k_1} \leq \liminf_{\alpha_3 \to 0} \mathbb{E} \sup_{0 \leq s \leq t} \left| \partial^{(\kappa)} X_s^{x'} - \partial^{(\kappa)} X_s^x \right|^{k_1}.$$

Therefore the same bounds in Lemma 4.3.2 and Theorem 4.3.3 apply if they are uniform with respect to small α_3; since γ is proportional to α_3 (for small γ), it follows that M in (4.3.3) and subsequently in α_t in both the proofs of Lemma 4.3.2 (in Lemma 4.3.1 and (4.3.11)) and Theorem 4.3.3 (equation (4.3.19)) blows up as $\alpha_3 \to 0$. This can be addressed by increasing m accordingly so that M is uniformly bounded.

4.8 Proofs of some auxiliary results

Just as in the case of globally Lipschitz coefficients in [114, Lemma 5.10], the regularity of an extended system and the harmonic property of the expectation (4.4.13) are required. These properties are established for our setting in the following.

Throughout the section, we assume $O = \mathbb{R}^n$ and b, σ are nonrandom functions. Moreover for the functions f, c, g, we assume nonrandomness and all of the presuppositions about them made in Theorem 4.4.2 regarding Assumption 9. In particular, $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}$, $c : [0, \infty) \times \mathbb{R}^n \to [0, \infty)$ and $g : \mathbb{R}^n \to \mathbb{R}$ are Borel functions satisfying that $f(t, \cdot), c(t, \cdot), g(\cdot)$ are continuous for every $t \in [0, T]$, $\int_0^T \sup_{x \in B_R} (|c(t, x)| + |f(t, x)|) dt < \infty$ for every $R > 0$ and such that for $h \in \{f, c, g\}$, $R > 0$, there exists $C \geq 0$, $0 < \bar{l} \leq 1$, Lyapunov functions $V^{s,T}$, locally bounded $\tilde{x}$ for which for any $s \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$|h(s + t, X_t^{s,x})| \leq C(1 + V^{s,T}(t, \tilde{x}(x)))^{\bar{l}} \tag{4.8.1a}$$

$$|h(s + t, y) - h(s + t, y')| \leq C|y - y'| \tag{4.8.1b}$$

$$V^{s+\tau,T}(0, \tilde{x}(X_\tau^{s,x}))^{\bar{l}} \leq C(1 + V^{s,T}(\tau, \tilde{x}(x))) \tag{4.8.1c}$$

for all $s \in [0, T]$, $t \leq T - s$, stopping times $\tau \leq T$, $x \in \mathbb{R}^n$ and $y, y' \in B_R$, where continuity of the underlying V_0 for the relevant Lyapunov functions have been used for (4.8.1b).

For any $s \geq 0$, $T > 0$, $x \in \mathbb{R}^n$, $x', x'' \in \mathbb{R}$, consider solutions $X_t^{s,x}$ to (4.2.1) appended with the corresponding $\mathbb{R}$-valued solutions $X_t^{(n+1),s,x'}$ and $X_t^{(n+2),s,x'}$ to

$$X_t^{(n+1),s,x'} = x' + \int_0^t c(s+r, X_r^{s,x})dr, \tag{4.8.2a}$$

$$X_t^{(n+2),s,x''} = x'' + \int_0^t f(s+r, X_r^{s,x})e^{-X_r^{(n+1),s,x'}}dr \tag{4.8.2b}$$

on $[0,T]$, denoted $\bar{X}_t^{s,y} = (X_t^{s,x}, X_t^{(n+1),s,x'}, X_t^{(n+2),s,x''})$, $y = (x, x', x'')$. Let $\bar{X}_t^{s,y}(I)$ be the corresponding Euler approximation analogous to (4.4.12) with I as in the beginning of Lemma 4.4.4.

Lemma 4.8.1. *Under the assumptions of this section, for every $R, T > 0$, it holds that*

$$\sup_{s\in[0,T]} \sup_{y\in B_R} \mathbb{P}\bigg(\sup_{t\in[0,T]} \big|\bar{X}_t^{s,y} - \bar{X}_t^{s,y}(I)\big| > \epsilon \bigg) \to 0$$

as $\sup_k t_{k+1} - t_k \to 0$.

Proof. For any $R' > 0$, let $R_X^{s,x}(I, R') \in \mathcal{F}$ denote the event

$$R_X^{s,x}(I,R') = \bigg\{ \sup_{t\in[0,T]} |X_t^{s,x}| \le R' \bigg\} \cap \bigg\{ \sup_{t\in[0,T]} |X_t^{s,x}(I)| \le R' \bigg\}.$$

For any $\epsilon, R' > 0$, it holds that

$$\begin{aligned}\mathbb{P}\bigg(\sup_{t\in[0,T]} \big|\bar{X}_t^{s,y} - \bar{X}_t^{s,y}(I)\big| > \epsilon \bigg) &\le \mathbb{P}\bigg(\sup_{t\in[0,T]} |X_t^{s,x}| > R' \bigg) + \mathbb{P}\bigg(\sup_{t\in[0,T]} |X_t^{s,x}(I)| > R' \bigg) \\ &\quad + \mathbb{P}\bigg(\sup_{t\in[0,T]} \big|\bar{X}_t^{s,y} - \bar{X}_t^{s,y}(I)\big| > \epsilon \,\bigg|\, R_X^{s,x}(I,R') \bigg).\end{aligned}$$

Fix $\epsilon' > 0$. For any $T, R > 0$, we may choose $R' = R^*$ so that, by Lemma 2.2 in [114], the sum of the first and second term on the right-hand side is bounded above by $\epsilon'/2$ uniformly in $s \in [0,T]$ and $x \in B_R$. For the last term on the right, note that by our assumptions on c, there exists locally bounded $\tilde{G} : \mathbb{R}^n \to [0,\infty)$ such that

$$\begin{aligned}&\sup_{t\in[0,T]} |c(s+t, X_t^{s,x}) - c(s+t, X_t^{s,x}(I))| \\ &\quad \le \sup_{t\in[0,T]} |X_t^{s,x} - X_t^{s,x}(I)|(\tilde{G}(X_t^{s,x}) + \tilde{G}(X_t^{s,x}(I)))\end{aligned} \tag{4.8.3}$$

and such that for $I_t := \max\{t_k : t \geq t_k\}$,

$$\begin{aligned}
&\sup_{t\in[0,T]} \left|c(s+t, X_t^{s,x}(I)) - c(s+t, X_{I_t}^{s,x}(I))\right| \\
&\quad \leq \sup_{t\in[0,T]} \left|X_t^{s,x}(I) - X_{I_t}^{s,x}(I)\right|(\tilde{G}(X_t^{s,x}(I)) + \tilde{G}(X_{I_t}^{s,x}(I))) \\
&\quad = \sup_{t\in[0,T]} \left|\int_{I_t}^{t} c(s+r, X_{I_t}^{s,x}(I))dr\right|(\tilde{G}(X_t^{s,x}(I)) + \tilde{G}(X_{I_t}^{s,x}(I))) \\
&\quad \leq 2 \sup_{t\in[0,T]} (t - I_t)(\tilde{G}(X_t^{s,x}(I)) + \tilde{G}(X_{I_t}^{s,x}(I)))
\end{aligned} \tag{4.8.4}$$

for all $s \in [0,T]$, $y = (x, x', x'') \in \mathbb{R}^{n+2}$, where we have used that c is uniformly bounded on $[0,2T] \times B_R$ by the continuity of the underlying V_0 for the Lyapunov function that forms an upper bound for c. By (4.8.3), it holds that

$$\begin{aligned}
&\left\{\sup_{t\in[0,T]} |X_t^{s,x} - X_t^{s,x}(I)| \leq \frac{\epsilon}{12\sqrt{3}T \sup_{z\in B_{R^*}} \tilde{G}(z)}\right\} \cap R_X^{s,x}(I, R^*) \\
&\quad \subset \left\{\sup_{t\in[0,T]} |c(s+t, X_t^{s,x}) - c(s+t, X_t^{s,x}(I))| \leq \frac{\epsilon}{6\sqrt{3}T}\right\} \cap R_X^{s,x}(I, R^*) \\
&\quad \subset \left\{\int_0^T |c(s+u, X_u^{s,x}) - c(s+u, X_u^{s,x}(I))| du \leq \frac{\epsilon}{6\sqrt{3}}\right\} \cap R_X^{s,x}(I, R^*)
\end{aligned}$$

and by (4.8.4), there exists $\delta^* > 0$ such that $\sup_k t_{k+1} - t_k \leq \delta^*$ implies

$$R_X^{s,x}(I, R^*) \subset \left\{\int_0^T \left|c(s+u, X_u^{s,x}(I)) - c(s+u, X_{I_u}^{s,x}(I))\right| du \leq \frac{\epsilon}{6\sqrt{3}}\right\}.$$

As a result, by Lemma 4.4.4 and our assumptions on b and σ, there exists $0 < \delta \leq \delta^*$ such that for I satisfying $\sup_{k\geq 0} t_{k+1} - t_k \leq \delta$, it holds that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,T]} \left|X_t^{(n+1),s,x'} - X_t^{(n+1),s,x'}(I)\right| > \frac{\epsilon}{3\sqrt{3}} \middle| R_X^{s,x}(I, R^*)\right) \\
&\quad \leq \mathbb{P}\left(\sup_{t\in[0,T]} |X_t^{s,x} - X_t^{s,x}(I)| > \frac{\epsilon}{12\sqrt{3}T \sup_{z\in B_{R^*}} \tilde{G}(z)}\right) \leq \frac{\epsilon'}{6}
\end{aligned} \tag{4.8.5}$$

for all $s \in [0,T]$ and $y = (x, x', x'') \in B_R \subset \mathbb{R}^{n+2}$. By a similar argument and using the above, (4.8.5) holds with $n+1$ replaced by $n+2$ and x' by x''. Together with Lemma 4.4.4, the lemma is proved. □

Next, the harmonic property (see [114, Definition 3.1]) of (4.4.13) is shown. Let $\bar{g}$ given by $\bar{g}(y) = x'' + g(x)e^{-x'}$ for all $y = (x, x', x'') \in \mathbb{R}^{n+2}$ and for $T > 0$, $s \in [0,T]$, let $\bar{v} : [0,\infty) \times \mathbb{R}^{n+2} \to \mathbb{R}$ be given by

$$\bar{v}(s,y) = \mathbb{E}[\bar{g}(\bar{X}^{s,y}_{T-s})] = x'' + v(s,x)e^{-x'} = \mathbb{E}[X^{(n+2),s,x''}_{T-s} + g(X^{s,x}_{T-s})e^{-X^{(n+1),s,x'}_{T-s}}]. \quad (4.8.6)$$

for v given in (4.4.13). In addition for a bounded subset $Q \subset (0,T) \times \mathbb{R}^{n+2}$, let τ be the stopping time

$$\tau := \inf\{u \geq 0 : (s+u, \bar{X}^{s,y}_u) \notin Q\}. \quad (4.8.7)$$

The next lemma establishes the equality $\bar{v}(s,y) = \mathbb{E}[\bar{v}(s+(\tau\wedge t), \bar{X}^{s,y}_{\tau\wedge t})]$ under our setting.

Lemma 4.8.2. *Under the assumption of this section, for any $T > 0$, any bounded subset $Q \subset (0,T) \times \mathbb{R}^{n+2}$, $(s,y) \in Q$, $t \in [0, T-s]$, it holds that*

$$\mathbb{E}[\bar{g}(\bar{X}^{s,y}_{T-s})] = \int\int \bar{g}(\bar{X}^{s+(\tau(\omega)\wedge t), \bar{X}^{s,y}_{\tau(\omega)\wedge t}(\omega)}_{T-s-(\tau(\omega)\wedge t)}(\omega'))d\mathbb{P}(\omega')d\mathbb{P}(\omega),$$

where τ is defined by (4.8.7).

Proof. For any $R, T > 0$, $t \in [0,T]$, $(s,y) \in Q$ with $y = (x, x', x'')$, by Theorem 2.13 in [114] together with Lemma 4.8.1, it holds for $\mathbb{P}$-a.a. ω that

$$\mathbb{E}[(\mathbb{1}_{B_R}\bar{g})(\bar{X}^{s,y}_{T-s})|\mathcal{F}_{\tau\wedge t}] = \int (\mathbb{1}_{B_R}\bar{g})(\bar{X}^{s+(\tau(\omega)\wedge t), \bar{X}^{s,y}_{\tau(\omega)\wedge t}(\omega)}_{T-(s+(\tau(\omega)\wedge t))}(\omega'))d\mathbb{P}(\omega'), \quad (4.8.8)$$

so that the right-hand side is $\mathcal{F}_{\tau\wedge t}$-measurable. Moreover for $\mathbb{P}$-a.a. ω, by (4.8.1), the absolute value of the integrand in the right-hand side is bounded independently of R as

$$\begin{aligned}
&(\mathbb{1}_{B_R}|\bar{g}|)(\bar{X}^{s+(\tau(\omega)\wedge t), \bar{X}^{s,y}_{\tau(\omega)\wedge t}(\omega)}_{T-(s+(\tau(\omega)\wedge t))}(\omega')) - |X^{(n+2),s,x''}_{\tau(\omega)\wedge t}(\omega)| \\
&\quad \leq \int_0^{T-s-(\tau(\omega)\wedge t)} |f(s+(\tau(\omega)\wedge t)+r, X^{s+(\tau(\omega)\wedge t), X^{s,x}_{\tau(\omega)\wedge t}(\omega)}_r(\omega'))|dr \\
&\qquad + |g(X^{s+(\tau(\omega)\wedge t), X^{s,x}_{\tau(\omega)\wedge t}(\omega)}_{T-(s+(\tau(\omega)\wedge t))}(\omega'))| \\
&\quad \leq C\bigg(\int_0^{T-s-(\tau(\omega)\wedge t)} (1 + V^{s+(\tau(\omega)\wedge t),T}(\omega', r, \tilde{x}(X^{s,x}_{\tau(\omega)\wedge t}(\omega))))^{\bar{l}} dr \\
&\qquad + (1 + V^{s+(\tau(\omega)\wedge t),T}(\omega', T-s-(\tau(\omega)\wedge t), \tilde{x}(X^{s,x}_{\tau(\omega)\wedge t}(\omega))))^{\bar{l}}\bigg).
\end{aligned} \quad (4.8.9)$$

Since $\bar{l}^{\text{th}}$-powers of Lyapunov functions are still Lyapunov functions (but with different auxiliary processes), the expectation in ω' of the right-hand side of which is bounded by

Theorem 2.4 in [101] and (4.8.1c) as in

$$\begin{aligned}
&\int \bigg(\int_0^{T-s-(\tau(\omega)\wedge t)} (1+V^{s+(\tau(\omega)\wedge t),T}(\omega', r, \tilde{x}(X^{s,x}_{\tau(\omega)\wedge t}(\omega))))^{\bar{l}} dr \\
&\quad + (1+V^{s+(\tau(\omega)\wedge t),T}(\omega', T-s-(\tau(\omega)\wedge t), \tilde{x}(X^{s,x}_{\tau(\omega)\wedge t}(\omega))))^{\bar{l}} \bigg) d\mathbb{P}(\omega') \\
&\quad \leq C \| e^{\int_0^T |\alpha_u^{s+(\tau(\omega)\wedge t),T}(\omega')| du} \|_{L^{\frac{p^{s+(\tau(\omega)\wedge t),T}}{p^{s+(\tau(\omega)\wedge t),T}-1}}(d\mathbb{P}(\omega'))} \\
&\qquad \cdot \int (1+V^{s+(\tau(\omega)\wedge t),T}(\omega', 0, \tilde{x}(X^{s,x}_{\tau(\omega)\wedge t}(\omega))))^{\bar{l}} d\mathbb{P}(\omega') \\
&\quad \leq C \bigg(1 + \int V^{s,T}(\omega', \tau(\omega)\wedge t, \tilde{x}(x)) d\mathbb{P}(\omega') \bigg) \\
&\quad \leq C(1+V^{s,T}(0,\tilde{x}(x))) \\
&\quad < \infty.
\end{aligned} \tag{4.8.10}$$

Therefore by dominated convergence, the right-hand side of (4.8.8) converges to the same expression but without $\mathbb{1}_{B_R}$ for $\mathbb{P}$-a.a. ω. Moreover, by (4.8.1) and Theorem 2.4 in [101],

$$\begin{aligned}
\mathbb{E}|X^{(n+2),s,x''}_{\tau\wedge t}| - |x''| &\leq \mathbb{E}\int_0^{\tau\wedge t} |f(s+r, X^{s,x''}_r)| dr \leq C\int_0^T \mathbb{E}[1+V^{s,T}(r,\tilde{x}(x''))] dr \\
&\leq C(1+V^{s,T}(0,\tilde{x}(x''))).
\end{aligned}$$

Consequently, together with (4.8.9), (4.8.10) and dominated convergence (in ω), it holds that

$$\begin{aligned}
&\int\int (\mathbb{1}_{B_R}\bar{g})(\bar{X}^{s+(\tau(\omega)\wedge t), \bar{X}^{s,y}_{\tau(\omega)\wedge t}(\omega)}_{T-(s+(\tau(\omega)\wedge t))}(\omega')) d\mathbb{P}(\omega') d\mathbb{P}(\omega) \\
&\quad \to \int\int \bar{g}(\bar{X}^{s+(\tau(\omega)\wedge t), \bar{X}^{s,y}_{\tau(\omega)\wedge t}(\omega)}_{T-(s+(\tau(\omega)\wedge t))}(\omega')) d\mathbb{P}(\omega') d\mathbb{P}(\omega)
\end{aligned}$$

as $R \to \infty$. On the other hand, by a similar argument as above, the expectation of the left-hand side of (4.8.8) has the limit

$$\mathbb{E}[\mathbb{E}[(\mathbb{1}_{B_R}\bar{g})(\bar{X}^{s,y}_{T-s})|\mathcal{F}_{\tau\wedge t}]] = \mathbb{E}[(\mathbb{1}_{B_R}\bar{g})(\bar{X}^{s,y}_{T-s})] \to \mathbb{E}[\bar{g}(\bar{X}^{s,y}_{T-s})]$$

as $R \to \infty$. $\square$

Bibliography

[1] S. A. Adelman and B. J. Garrison. Generalized Langevin theory for gas/solid processes: Dynamical solid models. *The Journal of Chemical Physics*, 65(9):3751–3761, 1976.

[2] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006. A hitchhiker's guide.

[3] H. AlRachid, L. Mones, and C. Ortner. Some remarks on preconditioning molecular dynamics. *SMAI J. Comput. Math.*, 4:57–80, 2018.

[4] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Stat. Comput.*, 18(4):343–373, 2008.

[5] L. Arnold, N. Sri Namachchivaya, and K. R. Schenk-Hoppé. Toward an understanding of stochastic Hopf bifurcation: a case study. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 6(11):1947–1975, 1996.

[6] A. D. Baczewski and S. D. Bond. Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel. *Journal of Chemical Physics*, 139(4):044107–044107, Jul 2013.

[7] J. Baker, P. Fearnhead, E. B. Fox, and C. Nemeth. Control variates for stochastic gradient MCMC. *Stat. Comput.*, 29(3):599–615, 2019.

[8] R. Bardenet, A. Doucet, and C. Holmes. On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.*, 18:Paper No. 47, 43, 2017.

[9] D. Belomestny, L. Iosipoi, E. Moulines, A. Naumov, and S. Samsonov. Variance reduction for Markov chains with application to MCMC. *Stat. Comput.*, 30(4):973–997, 2020.

[10] C. H. Bennett. Mass tensor molecular dynamics. *Journal of Computational Physics*, 19(3):267 – 279, 1975.

[11] A. Bertram and M. Grothaus. Essential m-dissipativity and hypocoercivity of Langevin dynamics with multiplicative noise. *J. Evol. Equ.*, 22(1):11, 2022.

[12] A. Beskos, G. Roberts, A. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stoch. Dyn.*, 8(3):319–350, 2008.

[13] A. Beskos and A. Stuart. MCMC methods for sampling function space. In *ICIAM 07—6th International Congress on Industrial and Applied Mathematics*, pages 337–364. Eur. Math. Soc., Zürich, 2009.

[14] M. Betancourt, S. Byrne, S. Livingstone, and M. Girolami. The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli*, 23(4A):2257–2298, 2017.

[15] R. N. Bhattacharya. 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.

[16] J. Bierkens, P. Fearnhead, and G. Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320, 2019.

[17] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2012. Anniversary edition [of MR1324786], With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.

[18] R. Biswas and D. R. Hamann. Simulated annealing of silicon atom clusters in Langevin molecular dynamics. *Phys. Rev. B*, 34:895–901, Jul 1986.

[19] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97:170201, Oct 2006.

[20] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov. *Fokker-Planck-Kolmogorov equations*, volume 207 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[21] F. Bolley, A. Guillin, and F. Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *M2AN Math. Model. Numer. Anal.*, 44(5):867–884, 2010.

[22] L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Rev.*, 60(2):223–311, 2018.

[23] N. Bou-Rabee and E. Vanden-Eijnden. Pathwise accuracy and ergodicity of metropolized integrators for SDEs. *Comm. Pure Appl. Math.*, 63(5):655–696, 2010.

[24] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.*, 113(522):855–867, 2018.

[25] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. Springer, New York, 2011.

[26] G. Bussi and M. Parrinello. Accurate sampling using Langevin dynamics. *Phys. Rev. E*, 75:056707, May 2007.

[27] J. Carrillo, S. Jin, L. Li, and Y. Zhu. A consensus-based global optimization method for high dimensional machine learning problems. 09 2019.

[28] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. An analytical framework for consensus-based global optimization method. *Math. Models Methods Appl. Sci.*, 28(6):1037–1066, 2018.

[29] P. Cattiaux, D. Chafaï, and A. Guillin. Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA Lat. Am. J. Probab. Math. Stat.*, 9(2):337–382, 2012.

[30] P. Cattiaux and L. Mesnager. Hypoelliptic non-homogeneous diffusions. *Probab. Theory Related Fields*, 123(4):453–483, 2002.

[31] M. Ceriotti. Generalized Langevin equation thermostats for ab initio molecular dynamics, 2014.

[32] M. Ceriotti, G. Bussi, and M. Parrinello. Langevin equation with colored noise for constant-temperature molecular dynamics simulations. *Phys. Rev. Lett.*, 102:020601, Jan 2009.

[33] M. Ceriotti, G. Bussi, and M. Parrinello. Colored-noise thermostats à la carte. *Journal of Chemical Theory and Computation*, 6(4):1170–1180, 2010.

[34] S. Cerrai. *Second order PDE's in finite and infinite dimension*, volume 1762 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001. A probabilistic approach.

[35] M. Chak. Regularity preservation in Kolmogorov equations for non-Lipschitz coefficients under Lyapunov conditions, 2022. arXiv: 2209.05436.

[36] M. Chak, N. Kantas, T. Lelièvre, and G. A. Pavliotis. Optimal friction matrix for underdamped Langevin sampling, 2021. arXiv: 2112.06844v1.

[37] M. Chak, N. Kantas, and G. A. Pavliotis. On the Generalised Langevin Equation for Simulated Annealing, 2021. arXiv: 2003.06448.

[38] M. Chaleyat-Maurel and D. Michel. Hypoellipticity theorems and conditional laws. *Z. Wahrsch. Verw. Gebiete*, 65(4):573–597, 1984.

[39] N. S. Chatterji, N. Flammarion, Y. Ma, Y. Ma, P. Bartlett, and M. I. Jordan. On the theory of variance reduction for stochastic gradient monte carlo. *ArXiv*, abs/1802.05431, 2018.

[40] X. Chen, S. Liu, R. Sun, and M. Hong. On the convergence of a class of adam-type algorithms for non-convex optimization. 2019. 7th International Conference on Learning Representations, ICLR 2019 ; Conference date: 06-05-2019 Through 09-05-2019.

[41] X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. In S. Bubeck, V. Perchet, and P. Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 300–323. PMLR, 06–09 Jul 2018.

[42] T.-S. Chiang, C.-R. Hwang, and S. J. Sheu. Diffusion for global optimization in $\mathbf{R}^n$. *SIAM J. Control Optim.*, 25(3):737–753, 1987.

[43] F. Conrad and M. Grothaus. Construction of N-particle Langevin dynamics for $H^{1,\infty}$-potentials via generalized Dirichlet forms. *Potential Anal.*, 28(3):261–282, 2008.

[44] S. Cox, M. Hutzenthaler, and A. Jentzen. Local lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations, 2013. arXiv:1309.5595. To appear in Mem. Amer. Math. Soc.

[45] D. Crisan, P. Dobson, B. Goddard, M. Ottobre, and I. Souttar. Poisson Equations with locally-lipschitz coefficients and Uniform in Time Averaging for Stochastic Differential Equations via Strong Exponential Stability, 2022.

[46] D. Crisan, P. Dobson, and M. Ottobre. Uniform in time estimates for the weak error of the Euler method for SDEs and a pathwise approach to derivative estimates for diffusion semigroups. *Trans. Amer. Math. Soc.*, 374(5):3289–3330, 2021.

[47] A. S. Dalalyan and L. Riou-Durand. On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.

[48] B. Delyon and Y. Hu. Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.*, 116(11):1660–1675, 2006.

[49] J. Diestel. Uniform integrability: an introduction. volume 23, pages 41–80 (1993). 1991. School on Measure Theory and Real Analysis (Grado, 1991).

[50] Z. Ding, Q. Li, J. Lu, and S. J. Wright. Random Coordinate Underdamped Langevin Monte Carlo, 2020. arXiv: 2010.11366.

[51] Z. Dong and X. Peng. Malliavin matrix of degenerate SDE and gradient estimate. *Electron. J. Probab.*, 19:no. 73, 26, 2014.

[52] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3):897–923, 2009.

[53] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[54] C. DuBois, A. Korattikara, M. Welling, and P. Smyth. Approximate slice sampling for bayesian posterior inference. In *Artificial Intelligence and Statistics*, pages 185–193. PMLR, 2014.

[55] A. B. Duncan, T. Lelièvre, and G. A. Pavliotis. Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.*, 163(3):457–491, 2016.

[56] A. B. Duncan, N. Nüsken, and G. A. Pavliotis. Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions. *J. Stat. Phys.*, 169(6):1098–1131, 2017.

[57] A. Durmus, A. Enfroy, Éric Moulines, and G. Stoltz. Uniform minorization condition and convergence bounds for discretizations of kinetic Langevin dynamics, 2021. arXiv: 2107.14542.

[58] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm, 2018. arXiv: 1605.01559.

[59] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

[60] R. Dwivedi, Y. Chen, M. J. Wainwright, and B. Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.*, 20:Paper No. 183, 42, 2019.

[61] A. Eberle, A. Guillin, and R. Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.*, 47(4):1982–2010, 2019.

[62] J.-P. Eckmann and M. Hairer. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.*, 212(1):105–164, 2000.

[63] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[64] J. C. M. Fok, B. Guo, and T. Tang. Combined Hermite spectral-finite difference method for the Fokker-Planck equation. *Math. Comp.*, 71(240):1497–1528, 2002.

[65] J. Foster, T. Lyons, and H. Oberhauser. The shifted ODE method for underdamped Langevin MCMC. *ArXiv*, abs/2101.03446, 2021.

[66] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, 3(4):391–412, 1999.

[67] A. Friedman. *Stochastic differential equations and applications. Vol. 1.* Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Probability and Mathematical Statistics, Vol. 28.

[68] S. Gadat and L. Miclo. Spectral decompositions and $\mathbb{L}^2$-operator norms of toy hypocoercive semi-groups. *Kinet. Relat. Models*, 6(2):317–372, 2013.

[69] S. Gadat and F. Panloup. Long time behaviour and stationary regime of memory gradient diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(2):564–601, 2014.

[70] X. Gao, M. Gurbuzbalaban, and L. Zhu. Breaking reversibility accelerates langevin dynamics for global non-convex optimization. *arXiv e-prints*, 12 2018. arXiv:1812.07725.

[71] S. B. Gelfand and S. K. Mitter. Recursive stochastic algorithms for global optimization in $\mathbf{R}^d$. *SIAM J. Control Optim.*, 29(5):999–1018, 1991.

[72] S. B. Gelfand and S. K. Mitter. Weak convergence of Markov chain sampling methods and annealing algorithms to diffusions. *J. Optim. Theory Appl.*, 68(3):483–498, 1991.

[73] S. Gemam and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence - PAMI*, PAMI-6:721–741, 1984.

[74] S. Geman and C.-R. Hwang. Diffusions for global optimization. *SIAM J. Control Optim.*, 24(5):1031–1043, 1986.

[75] E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson. Global convergence of the heavy-ball method for convex optimization. In *2015 European Control Conference (ECC)*, pages 310–315, 2015.

[76] S. Ghadimi and G. Lan. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Mathematical Programming*, 156(1):59–99, 2016.

[77] B. Gidas. Global optimization via the Langevin equation. In *1985 24th IEEE Conference on Decision and Control*, pages 774 – 778, Dec 1985.

[78] B. Gidas. Nonstationary Markov chains and convergence of the annealing algorithm. *J. Statist. Phys.*, 39(1-2):73–131, 1985.

[79] P. J. Green, K. Ł atuszyński, M. Pereyra, and C. P. Robert. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Stat. Comput.*, 25(4):835–862, 2015.

[80] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[81] A. Guillin and P. Monmarché. Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electron. Commun. Probab.*, 21:Paper No. 74, 14, 2016.

[82] A. Guillin and F.-Y. Wang. Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality. *J. Differential Equations*, 253(1):20–40, 2012.

[83] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[84] B.-Y. Guo. *Spectral methods and their applications.* World Scientific Publishing Co., Inc., River Edge, NJ, 1998.

[85] M. Hairer, M. Hutzenthaler, and A. Jentzen. Loss of regularity for Kolmogorov equations. *Ann. Probab.*, 43(2):468–527, 2015.

[86] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.*, 17(5-6):1657–1706, 2007.

[87] M. Hairer, A. M. Stuart, and J. Voss. Sampling conditioned hypoelliptic diffusions. *Ann. Appl. Probab.*, 21(2):669–698, 2011.

[88] M. Hairer, A. M. Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs arising in path sampling. I. The Gaussian case. *Commun. Math. Sci.*, 3(4):587–603, 2005.

[89] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[90] Y. He, K. Balasubramanian, and M. A. Erdogdu. On the Ergodicity, Bias and Asymptotic Normality of Randomized Midpoint Sampling Method, 2020. arXiv: 2011.03176.

[91] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[92] R. Holley and D. Stroock. Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.*, 115(4):553–569, 1988.

[93] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[94] R. Höpfner, E. Löcherbach, and M. Thieullen. Strongly degenerate time inhomogeneous SDEs: densities and support properties. Application to Hodgkin-Huxley type systems. *Bernoulli*, 23(4A):2587–2616, 2017.

[95] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[96] A. M. Horowitz. The second order Langevin equation and numerical simulations. *Nuclear Physics B*, 280:510–522, 1987.

[97] A. M. Horowitz. A generalized guided Monte Carlo algorithm. *Physics Letters B*, 268(2):247–252, 1991.

[98] S. Hottovy, A. McDaniel, G. Volpe, and J. Wehr. The Smoluchowski-Kramers limit of stochastic differential equations with arbitrary state-dependent friction. *Comm. Math. Phys.*, 336(3):1259–1283, 2015.

[99] A. Hudde, M. Hutzenthaler, A. Jentzen, and S. Mazzonetto. On the Itô-Alekseev-Gröbner formula for stochastic differential equations, 2018. arXiv: 1812.09857. To appear in Ann. Inst. Henri Poincaré Probab. Stat.

[100] A. Hudde, M. Hutzenthaler, and S. Mazzonetto. On moments and strong local Hölder regularity of solutions of stochastic differential equations and of their spatial derivative processes, 2019.

[101] A. Hudde, M. Hutzenthaler, and S. Mazzonetto. A stochastic Gronwall inequality and applications to moments, strong completeness, strong local Lipschitz continuity, and perturbations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):603–626, 2021.

[102] M. Hutzenthaler and A. Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Mem. Amer. Math. Soc.*, 236(1112):v+99, 2015.

[103] M. Hutzenthaler and A. Jentzen. On a perturbation theory and on strong convergence rates for stochastic ordinary and partial differential equations with nonglobally monotone coefficients. *Ann. Probab.*, 48(1):53–93, 2020.

[104] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 467(2130):1563–1576, 2011.

[105] M. Hutzenthaler, A. Jentzen, and X. Wang. Exponential integrability properties of numerical approximation processes for nonlinear stochastic differential equations. *Math. Comp.*, 87(311):1353–1413, 2018.

[106] C.-R. Hwang. Laplace's method revisited: weak convergence of probability measures. *Ann. Probab.*, 8(6):1177–1182, 1980.

[107] C.-R. Hwang and S. J. Sheu. Large-time behavior of perturbed diffusion Markov processes with applications to the second eigenvalue problem for Fokker-Planck operators and simulated annealing. *Acta Appl. Math.*, 19(3):253–295, 1990.

[108] H. Ishii. On the equivalence of two notions of weak solutions, viscosity solutions and distribution solutions. *Funkcial. Ekvac.*, 38(1):101–120, 1995.

[109] A. Jentzen, B. Kuckuck, T. Müller-Gronbach, and L. Yaroslavtseva. Counterexamples to local lipschitz and local hölder continuity with respect to the initial values for additive noise driven stochastic differential equations with smooth drift coefficient functions with at most polynomially growing derivatives. *Discrete and Continuous Dynamical Systems - B*, 0:–, 2021.

[110] A. Kavalur, V. Guduguntla, and W. K. Kim. Effects of Langevin friction and time steps in the molecular dynamics simulation of nanoindentation. *Molecular Simulation*, 46(12):911–922, 2020.

[111] R. Khasminskii. *Stochastic stability of differential equations*, volume 66 of *Stochastic Modelling and Applied Probability.* Springer, Heidelberg, second edition, 2012. With contributions by G. N. Milstein and M. B. Nevelson.

[112] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1992.

[113] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov processes*, volume 345 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2012. Time symmetry and martingale approximation.

[114] N. V. Krylov. On Kolmogorov's equations for finite-dimensional diffusions. In *Stochastic PDE's and Kolmogorov equations in infinite dimensions (Cetraro, 1998)*, volume 1715 of *Lecture Notes in Math.*, pages 1–63. Springer, Berlin, 1999.

[115] H. J. Kushner. Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via Monte Carlo. *SIAM J. Appl. Math.*, 47(1):169–185, 1987.

[116] H. Lei, N. A. Baker, and X. Li. Data-driven parameterization of the generalized Langevin equation. *Proc. Natl. Acad. Sci. USA*, 113(50):14183–14188, 2016.

[117] B. Leimkuhler and C. Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. Express. AMRX*, (1):34–56, 2013.

[118] B. Leimkuhler and C. Matthews. Robust and efficient configurational molecular sampling via langevin dynamics. *The Journal of Chemical Physics*, 138(17):174102, 2013.

[119] B. Leimkuhler and C. Matthews. *Molecular dynamics*, volume 39 of *Interdisciplinary Applied Mathematics*. Springer, Cham, 2015. With deterministic and stochastic numerical methods.

[120] B. Leimkuhler, C. Matthews, and G. Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.*, 36(1):13–79, 2016.

[121] B. Leimkuhler and M. Sachs. Efficient Numerical Algorithms for the Generalized Langevin Equation. *arXiv e-prints*, page arXiv:2012.04245, Dec. 2020.

[122] T. Lelièvre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.*, 152(2):237–274, 2013.

[123] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations*. Imperial College Press, London, 2010. A mathematical perspective.

[124] T. Lelièvre and G. Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.*, 25:681–880, 2016.

[125] C. Li, C. Chen, D. Carlson, and L. Carin. Preconditioned stochastic gradient langevin dynamics for deep neural networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1788–1794. AAAI Press, 2016.

[126] X.-M. Li and M. Scheutzow. Lack of strong completeness for stochastic flows. *Ann. Probab.*, 39(4):1407–1421, 2011.

[127] P.-L. Lions. Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. II. Viscosity solutions and uniqueness. *Comm. Partial Differential Equations*, 8(11):1229–1276, 1983.

[128] L. Lorenzi and M. Bertoldi. *Analytical methods for Markov semigroups*, volume 283 of *Pure and Applied Mathematics (Boca Raton)*. Chapman & Hall/CRC, Boca Raton, FL, 2007.

[129] Y.-A. Ma, N. S. Chatterji, X. Cheng, N. Flammarion, P. L. Bartlett, and M. I. Jordan. Is there an analog of Nesterov acceleration for gradient-based MCMC? *Bernoulli*, 27(3):1942–1992, 2021.

[130] Y.-A. Ma, Y. Chen, C. Jin, N. Flammarion, and M. I. Jordan. Sampling can be faster than optimization. *Proc. Natl. Acad. Sci. USA*, 116(42):20881–20885, 2019.

[131] D. Márquez. Convergence rates for annealing diffusion processes. *Ann. Appl. Probab.*, 7(4):1118–1139, 1997.

[132] J. C. Mattingly and A. M. Stuart. Geometric ergodicity of some hypo-elliptic diffusions for particle motions. volume 8, pages 199–214. 2002. Inhomogeneous random systems (Cergy-Pontoise, 2001).

[133] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[134] G. Menz and A. Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.*, 42(5):1809–1884, 2014.

[135] G. Metafune, D. Pallara, and E. Priola. Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures. *J. Funct. Anal.*, 196(1):40–60, 2002.

[136] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[137] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Adv. in Appl. Probab.*, 25(3):487–517, 1993.

[138] L. Miclo. Recuit simulé sur $\mathbf{R}^n$. Étude de l'évolution de l'énergie libre. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(2):235–266, 1992.

[139] P. Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *Probab. Theory Related Fields*, 172(3-4):1215–1248, 2018.

[140] P. Monmarché. Generalized Γ calculus and application to interacting particles on a graph. *Potential Anal.*, 50(3):439–466, 2019.

[141] P. Monmarché. High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion. *Electron. J. Stat.*, 15(2):4117–4166, 2021.

[142] P. Monmarché. Almost sure contraction for diffusions on $\setminus^d$. Application to generalised Langevin diffusions, 2021. arXiv: 2009.10828.

[143] W. Mou, Y.-A. Ma, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan. High-order langevin diffusion yields an accelerated mcmc algorithm. *Journal of Machine Learning Research*, 22(42):1–41, 2021.

[144] W. Mou, Y.-A. Ma, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan. High-order Langevin diffusion yields an accelerated MCMC algorithm. *J. Mach. Learn. Res.*, 22:Paper No. 42, 41, 2021.

[145] M. Nava, M. Ceriotti, C. Dryzun, and M. Parrinello. Evaluating functions of positive-definite matrices using colored-noise thermostats. *Phys. Rev. E*, 89:023302, Feb 2014.

[146] R. Neal. Bayesian learning via stochastic dynamics. In S. Hanson, J. Cowan, and C. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5. Morgan-Kaufmann, 1993.

[147] R. M. Neal. Slice sampling. *Ann. Statist.*, 31(3):705–767, 2003. With discussions and a rejoinder by the author.

[148] R. M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 113–162. CRC Press, Boca Raton, FL, 2011.

[149] C. Nemeth and P. Fearnhead. Stochastic Gradient Markov Chain Monte Carlo. *J. Amer. Statist. Assoc.*, 116(533):433–450, 2021.

[150] Y. Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, Cham, 2018. Second edition of [MR2142598].

[151] M. Ottobre and G. A. Pavliotis. Asymptotic analysis for the generalized Langevin equation. *Nonlinearity*, 24(5):1629–1653, 2011.

[152] M. Ottobre, G. A. Pavliotis, and K. Pravda-Starov. Exponential return to equilibrium for hypoelliptic quadratic systems. *J. Funct. Anal.*, 262(9):4000–4039, 2012.

[153] A. B. Owen. Statistically efficient thinning of a Markov chain sampler. *J. Comput. Graph. Statist.*, 26(3):738–744, 2017.

[154] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.

[155] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. III. *Ann. Probab.*, 33(3):1111–1133, 2005.

[156] S. Patterson and Y. W. Teh. Stochastic gradient Riemannian Langevin dynamics on the probability simplex. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3102–3110. Curran Associates, Inc., 2013.

[157] G. A. Pavliotis. Asymptotic analysis of the Green-Kubo formula. *IMA J. Appl. Math.*, 75(6):951–967, 2010.

[158] G. A. Pavliotis. *Stochastic processes and applications*, volume 60 of *Texts in Applied Mathematics*. Springer, New York, 2014. Diffusion processes, the Fokker-Planck and Langevin equations.

[159] G. A. Pavliotis, G. Stoltz, and U. Vaes. Scaling limits for the generalized Langevin equation. *J. Nonlinear Sci.*, 31(1):Paper No. 8, 58, 2021.

[160] M. Pelletier. Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1):10–44, 1998.

[161] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci.*, 27(1):183–204, 2017.

[162] B. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[163] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[164] P. Protter, L. Qiu, and J. S. Martin. Asymptotic error distribution for the Euler scheme with locally Lipschitz coefficients. *Stochastic Process. Appl.*, 130(4):2296–2311, 2020.

[165] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[166] S. Redon, G. Stoltz, and Z. Trstanova. Error analysis of modified Langevin dynamics. *J. Stat. Phys.*, 164(4):735–771, 2016.

[167] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[168] J. Roussel and G. Stoltz. Spectral methods for Langevin dynamics and associated error estimates. *ESAIM Math. Model. Numer. Anal.*, 52(3):1051–1083, 2018.

[169] G. Royer. A remark on simulated annealing of diffusion processes. *SIAM J. Control Optim.*, 27(6):1403–1408, 1989.

[170] S. Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.

[171] M. Sachs. *The generalised Langevin equation: asymptotic properties and numerical analysis*. PhD thesis, The University of Edinburgh, 2017.

[172] M. Sachs, B. Leimkuhler, and V. Danos. Langevin dynamics with variable coefficients and nonconservative forces: From stationary states to numerical methods. *Entropy*, 19(12), 2017.

[173] J. M. Sanz-Serna and K. Zygalakis. Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations. *ArXiv*, abs/2104.12384, 2021.

[174] A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès, and M. Caffarel. An efficient sampling algorithm for variational monte carlo. *The Journal of Chemical Physics*, 125(11):114105, 2006.

[175] M. Scheutzow. A stochastic Gronwall lemma. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 16(2):1350019, 4, 2013.

[176] A. Schlichting. The Eyring-Kramers formula for Poincaré and logarithmic Sobolev inequalities. 2012.

[177] R. Shen and Y. T. Lee. The randomized midpoint method for log-concave sampling. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[178] R. D. Skeel and C. Hartmann. Choice of Damping Coefficient in Langevin Dynamics, 2021. arXiv: 2106.11597.

[179] H. Song, I. Triguero, and E. Özcan. A review on the self and dual interactions between machine learning and optimisation. *Progress in Artificial Intelligence*, 8(2):143–165, 2019.

[180] L. F. South, C. J. Oates, A. Mira, and C. Drovandi. Regularised Zero-Variance Control Variates for High-Dimensional Variance Reduction, 2020. arXiv: 1811.05073.

[181] G. Stoltz and Z. Trstanova. Langevin dynamics with general kinetic energies. *Multiscale Model. Simul.*, 16(2):777–806, 2018.

[182] D. W. Stroock and S. R. S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory*, pages 333–359, 1972.

[183] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *J. Mach. Learn. Res.*, 17:Paper No. 153, 43, 2016.

[184] Y. Sun and A. Garcia. Interactive diffusions for global optimization. *J. Optim. Theory Appl.*, 163(2):491–509, 2014.

[185] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

[186] S. Taniguchi. Applications of Malliavin's calculus to time-dependent systems of heat equations. *Osaka J. Math.*, 22(2):307–320, 1985.

[187] J. Teichmann. Calculating the Greeks by cubature formulae. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 462(2066):647–670, 2006.

[188] Z. Trstanova and S. Redon. Estimating the speed-up of adaptively restrained Langevin dynamics. *J. Comput. Phys.*, 336:412–428, 2017.

[189] P. Vanetti, A. Bouchard-Côté, G. Deligiannidis, and A. Doucet. Piecewise-Deterministic Markov Chain Monte Carlo, 2018. arXiv: 1707.05296.

[190] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[191] S. J. Vollmer, K. C. Zygalakis, and Y. W. Teh. Exploration of the (non-)asymptotic bias and variance of stochastic gradient Langevin dynamics. *J. Mach. Learn. Res.*, 17:Paper No. 159, 45, 2016.

[192] X. Wang and Y. Zhao. Weak approximations of nonlinear SDEs with non-globally Lipschitz continuous coefficients, 2021.

[193] M. Welling and Y. W. Teh. Bayesian Learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, page 681–688, Madison, WI, USA, 2011. Omnipress.

[194] L. Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.*, 91(2):205–238, 2001.

[195] X. Wu, B. R. Brooks, and E. Vanden-Eijnden. Self-guided Langevin dynamics via generalized Langevin equation. *J Comput Chem*, 37(6):595–601, Mar 2016.

[196] J. Yang, G. O. Roberts, and J. S. Rosenthal. Optimal scaling of random-walk metropolis algorithms on general target distributions. *Stochastic Process. Appl.*, 130(10):6094–6132, 2020.

[197] D. Zou and Q. Gu. On the Convergence of Hamiltonian Monte Carlo with Stochastic Gradients. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 13012–13022. PMLR, 18–24 Jul 2021.