

Underdamped Langevin MCMC with third order convergence

Maximilian Scott

MS4314@BATH.AC.UK

Dáire O’Kane

DOK29@BATH.AC.UK

Andraž Jelinčič

AJ2382@BATH.AC.UK

James Foster

JMF68@BATH.AC.UK

*Department of Mathematical Sciences
University of Bath, UK*

Editor: Kevin Jamieson

Abstract

In this paper, we propose a new numerical method for the underdamped Langevin diffusion (ULD) and present a non-asymptotic analysis of its sampling error in the 2-Wasserstein distance when the d -dimensional target distribution $p(x) \propto e^{-f(x)}$ is strongly log-concave and has varying degrees of smoothness. Precisely, under the assumptions that the gradient and Hessian of f are Lipschitz continuous, our algorithm achieves a 2-Wasserstein error of ε in $\mathcal{O}(\sqrt{d}/\varepsilon)$ and $\mathcal{O}(\sqrt{d}/\sqrt{\varepsilon})$ steps respectively. Therefore, our algorithm has a similar complexity as other popular Langevin MCMC algorithms under matching assumptions. However, if we additionally assume that the third derivative of f is Lipschitz continuous, then our algorithm achieves a 2-Wasserstein error of ε in $\mathcal{O}(\sqrt{d}/\varepsilon^{\frac{1}{3}})$ steps. To the best of our knowledge, this is the first gradient-only method for ULD with third order convergence. To support our theory, we perform Bayesian logistic regression across a range of real-world datasets, where our algorithm achieves competitive performance compared to an existing underdamped Langevin MCMC algorithm and the popular No U-Turn Sampler (NUTS).

Keywords: Langevin dynamics, Markov Chain Monte Carlo, High order convergence

1. Introduction

In a variety of applications ranging from statistical inference (Ma et al., 2015; Mandt et al., 2017; Durmus and Moulines, 2019) to molecular dynamics (Pastor et al., 1988; Pastor, 1994; Allen and Tildesley, 2017), we are often interested in sampling from a target distribution

$$p(x) \propto e^{-f(x)}, \quad (1)$$

where $x \in \mathbb{R}^d$ and $f : \mathbb{R}^d \rightarrow \mathbb{R}$ denotes a scalar potential. In practice, we typically only know this target distribution up to a normalisation constant – which becomes challenging to compute when the dimension d is large. So instead, we can generate samples by constructing and simulating a Markov chain whose marginal distribution becomes close to the target (1). This methodology for sampling from p is known as Markov Chain Monte Carlo (or MCMC).

In this paper, we consider MCMC algorithms obtained by simulating the underdamped Langevin diffusion (ULD), given by the following stochastic differential equation (SDE):

$$\begin{aligned} dx_t &= v_t dt, \\ dv_t &= -\gamma v_t dt - u \nabla f(x_t) dt + \sqrt{2\gamma u} dW_t, \end{aligned} \tag{2}$$

where $x, v \in \mathbb{R}^d$ represent the position and momentum of a particle, $\gamma > 0$ denotes a friction coefficient, $u > 0$ and W is a d -dimensional Brownian motion. Within the context of MCMC, one can consider γ and u as tuneable hyperparameters with the friction γ corresponding to the amount of partial momentum refreshment in Hamiltonian Monte Carlo (Riou-Durand and Vogrinc, 2022; Riou-Durand et al., 2023). Under mild assumptions on the scalar potential f , it is known that the SDE (2) admits a unique strong solution which is ergodic with respect to the Gibbs measure with Hamiltonian $H(x, v) = f(x) + \frac{1}{2u} \|v\|^2$ (Pavliotis, 2014, Proposition 6.1). In other words, ULD has a stationary measure π on \mathbb{R}^{2d} with density proportional to $\exp(-H(x, v))$. Therefore, the distribution of x_t will converge to the target distribution (1) and thus simulating ULD will allow us to generate samples. In general, the SDE (2) which governs ULD does not admit a closed-form solution, and so it is necessary to develop appropriate numerical methods for approximating these samples.

A variety of approximations have previously been proposed for ULD. Some noteworthy examples include the unadjusted and Metropolis-adjusted OBABO schemes (Bussi and Parrinello, 2006; Monmarché, 2021; Song and Tan, 2021; Riou-Durand and Vogrinc, 2022), the UBU method (Álamo and Sanz-Serna, 2019; Sanz-Serna and Zygalkakis, 2021), the ALUM methods (Hu et al., 2021), the randomised midpoint method (Shen and Lee, 2019), the BAOAB method (Leimkuhler et al., 2016) and a Strang splitting (Buckwar et al., 2019). In addition, there exist a variety of methods for improving the accuracy and robustness of underdamped Langevin MCMC algorithms. These include using a preconditioning matrix (Goodman and Weare, 2010; Leimkuhler et al., 2018; AlRachid et al., 2018), adaptive step sizes (Foster and Jelinčič, 2024; Leroy et al., 2024; Leimkuhler et al., 2025) or unbiased estimators based on Multilevel Monte Carlo techniques (Giles, 2015; Chada et al., 2024).

In this paper, we introduce a practical method for approximating ULD called QUICSORT¹, which is derived using a piecewise linear discretisation of the underlying Brownian motion.

Definition 1 (QUICSORT) *Let $\{t_n\}_{n \geq 0}$ be a sequence of times with $t_0 = 0$, $t_{n+1} > t_n$ and $h_n = t_{n+1} - t_n$. We construct a numerical solution $\{(X_n, V_n)\}_{n \geq 0}$ for the SDE (2) by choosing some initial $(X_0, V_0) \sim \pi_0$ and, for each $n \geq 0$, defining (X_{n+1}, V_{n+1}) as follows:*

$$\begin{aligned} V_n^{(1)} &:= V_n + \sigma(H_n + 6K_n), \\ X_n^{(1)} &:= X_n + \frac{1 - e^{-\lambda - \gamma h_n}}{\gamma} V_n^{(1)} + \frac{e^{-\lambda - \gamma h_n} + \lambda_- \gamma h_n - 1}{\gamma^2 h_n} C_n, \\ X_n^{(2)} &:= X_n + \frac{1 - e^{-\lambda + \gamma h_n}}{\gamma} V_n^{(1)} - \frac{1 - e^{-\frac{1}{3}\gamma h_n}}{\gamma} u \nabla f(X_n^{(1)}) h_n + \frac{e^{-\lambda + \gamma h_n} + \lambda_+ \gamma h_n - 1}{\gamma^2 h_n} C_n, \\ V_n^{(2)} &:= e^{-\gamma h_n} V_n^{(1)} - \frac{1}{2} e^{-\lambda + \gamma h_n} u \nabla f(X_n^{(1)}) h_n - \frac{1}{2} e^{-\lambda - \gamma h_n} u \nabla f(X_n^{(2)}) h_n + \frac{1 - e^{-\gamma h_n}}{\gamma h_n} C_n, \end{aligned}$$

1. **QU**adrature **I**nspired and **C**ontractive **S**hifted **O**DE with **R**unge-Kutta **T**hree

$$\begin{aligned}
 X_{n+1} := & X_n - \frac{1 - e^{-\lambda_+ \gamma h_n}}{2\gamma} u \nabla f(X_n^{(1)}) h_n - \frac{1 - e^{-\lambda_- \gamma h_n}}{2\gamma} u \nabla f(X_n^{(2)}) h_n \\
 & + \frac{1 - e^{-\gamma h_n}}{\gamma} V_n^{(1)} + \frac{e^{-\gamma h_n} + \gamma h_n - 1}{\gamma^2 h_n} C_n,
 \end{aligned} \tag{3}$$

$$V_{n+1} := V_n^{(2)} - \sigma(H_n - 6K_n).$$

where $\sigma = \sqrt{2\gamma u}$, $\lambda_+ = \frac{3+\sqrt{3}}{6}$, $\lambda_- = \frac{3-\sqrt{3}}{6}$ and $C_n = \sigma(W_n - 12K_n)$. The vectors $\{W_n\}_{n \geq 0}$, $\{H_n\}_{n \geq 0}$ and $\{K_n\}_{n \geq 0}$ are independent and Gaussian distributed with $W_n \sim \mathcal{N}(0, h_n I_d)$, $H_n \sim \mathcal{N}(0, \frac{1}{12} h_n I_d)$ and $K_n \sim \mathcal{N}(0, \frac{1}{720} h_n I_d)$ for all $n \geq 0$. These random vectors can be viewed as the first three coefficients in a polynomial expansion of the Brownian motion W (see Foster et al. (2020)).

This approach can be seen as a specific discretisation of the ‘‘Shifted ODE’’ introduced in Foster et al. (2021). Whilst under certain convexity and smoothness assumptions on f , the shifted ODE approximation achieves third order convergence guarantees that hold for all $n \geq 0$, this convergence does not automatically extend to practical numerical discretisations (e.g. the SORT and SOFA methods originally proposed for discretising the shifted ODE). Therefore, our work directly improves upon Foster et al. (2021) since we both theoretically guarantee third order convergence for QUICSORT as well as demonstrate it numerically.

In the recent literature, it has become standard practice to understand a method’s efficacy through a non-asymptotic error analysis. More precisely, by finding an upper bound for the 2-Wasserstein distance between the target distribution (1) and the distribution of X_n , the position variable at the n -th iteration, we can establish a mixing rate for the different numerical methods. Quantitatively, we compare the number of steps n required to reach an error of $W_2(X_n, e^{-f}) \leq \varepsilon$. The main theoretical contribution of this paper is that we show that the QUICSORT method achieves a 2-Wasserstein error of ε in $\mathcal{O}(\sqrt{d} \varepsilon^{-\frac{1}{3}})$ steps under certain smoothness and strong convexity assumption on f . Our central result is as follows:

Theorem 2 (QUICSORT non-asymptotic convergence) *Let $\{(X_n, V_n)\}_{n \geq 0}$ denote the QUICSORT approximation (3) where $h_n = h$ for some fixed $h \in (0, 1]$. Suppose that f is three times differentiable, m -strongly convex and its gradient, Hessian and third derivative are all Lipschitz continuous (with respective constants M_1, M_2, M_3) and that $\gamma \geq 2\sqrt{uM_1}$. Assume further that X_0 has finite second moments and $V_0 \sim \mathcal{N}(0, uI_d)$. Then there exists constants $\beta, C > 0$ (not depending on d and h) such that for all $n \geq 0$, the 2-Wasserstein distance between the law of X_n and the target distribution $p(x) \propto e^{-f(x)}$ can be bounded as*

$$W_2(\text{Law}(X_n), p) \leq ce^{-n\beta h} W_2(\text{Law}(X_0), p) + Cd^{1.5}h^3. \tag{4}$$

In general, MCMC algorithms are only concerned with approximating statistics of the target distribution, and so only weak convergence is needed. Fortunately, by obtaining bounds on the strong² error of QUICSORT, we also obtain bounds for the 2-Wasserstein error (as given in (4)). However, as QUICSORT has state-of-the-art strong convergence properties, it may also have advantages over purely ‘weak’ MCMC algorithms. Firstly, it is straightforward to numerically estimate the strong convergence of QUICSORT, which could help quantify

2. Throughout the paper, the terminology ‘‘strong’’ error refers to the \mathbb{L}_2 (or root mean square) error.

its efficacy or be a diagnostic for determining when the step size is too large. Similarly, we expect high order strong convergence to be helpful when more advanced methodologies are applied, such as adaptive timestepping (Jelinčič et al., 2024) or unbiased estimators based on either synchronous coupling or Multilevel Monte Carlo techniques (Chada et al., 2024).

In Table 1, we compare the mixing rate of QUICSORT with other approximations of ULD. For proofs of their convergence rates see Monmarché (2021), Sanz-Serna and Zygalakis (2021, Sections 6.1 and 6.2), Paulin and Whalley (2024), Dalalyan and Riou-Durand (2020) and Hu et al. (2021) respectively.

Numerical method	Smoothness assumptions on the strongly convex f	Number of steps n to achieve an error of $W_2(Law(X_n), p) \leq \varepsilon$
QUICSORT (ours)	Lipschitz Gradient + Lipschitz Hessian + Lipschitz third derivative	$\mathcal{O}(\sqrt{d}/\varepsilon)$ $\mathcal{O}(\sqrt{d}/\sqrt{\varepsilon})$ $\mathcal{O}(\sqrt{d}/\varepsilon^{\frac{1}{3}})$
OBABO splitting	Lipschitz Gradient + Lipschitz Hessian	$\mathcal{O}(\sqrt{d}/\varepsilon)$ $\mathcal{O}(\sqrt{d}/\sqrt{\varepsilon})$
Exponential Euler	Lipschitz Gradient	$\mathcal{O}(\sqrt{d}/\varepsilon)$
UBU splitting	Lipschitz Gradient + Lipschitz Hessian + Strongly Lipschitz Hessian	$\mathcal{O}(\sqrt{d}/\varepsilon)$ $\mathcal{O}(\sqrt{d}/\sqrt{\varepsilon})$ $\mathcal{O}(\sqrt[4]{d}/\sqrt{\varepsilon})$
ALUM	Lipschitz Gradient	$\mathcal{O}(\sqrt[3]{d}/\varepsilon^{\frac{2}{3}})$

Table 1: Summary of complexities for ULD methods (with respect to 2-Wasserstein error).

We can numerically verify the mixing rates for most of these methods by estimating their strong error at a large time T . This involves comparing the numerical solution obtained by the method using both large and small step sizes. To ensure that these two numerical solutions are close, we will apply methods using the same underlying Brownian motion. The estimated strong convergence rates for some of the numerical methods in Table 1 are presented in Figure 1. We give a more detailed description of our experiments in Section 4. To help facilitate applications, our QUICSORT method has been implemented in DiffraX (Kidger, 2021), which is a high performance open-source software package for ODE and SDE simulation (see docs.kidger.site/diffrax/examples/underdamped_langevin_example).

The paper is organised as follows. In Section 2 we establish our key notation, definitions and assumptions. We also provide an outline for the derivation of the QUICSORT method. In Section 3, we present our main results (first, second and third order convergence rates for the QUICSORT method) and describe our strategy of proof. In Section 4, we provide experiments demonstrating the convergence and efficacy of the QUICSORT method as an MCMC algorithm for performing Bayesian logistic regression on real-world data. Finally, the technical details of our non-asymptotic error analysis can be found in the appendix.

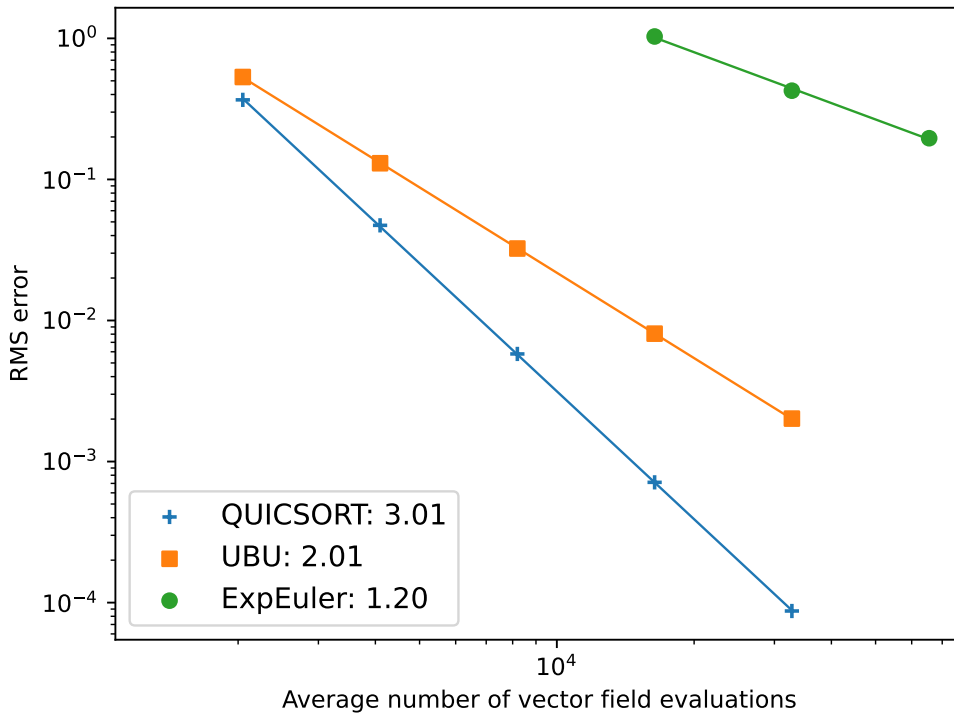


Figure 1: Empirical validation of the complexities in Table 1. Here, the numerical methods are applied to simulate underdamped Langevin dynamics where the potential f is obtained from a Bayesian logistic regression that is applied to the Titanic passenger survival dataset (Hendricks, 2015). We compute the Monte Carlo estimators for the strong error at $T = 100$.

2. Notation, definitions and derivation of the QUICSORT method

In this section, we present some of the notation, definitions and assumptions for the paper. We then provide an informal derivation of the QUICSORT method (3).

Throughout the paper, we use $\|\cdot\|_2$ to denote the standard Euclidean norm on \mathbb{R}^d and \mathbb{R}^{2d} . Along with the Euclidean norm, we will use the usual inner product $\langle \cdot, \cdot \rangle$ with $\langle v, v \rangle = \|v\|_2^2$.

We shall now briefly recall the notions of tensor and operator norms. For $k \geq 1$, a k -tensor on \mathbb{R}^d is an element of the d^k -dimensional space $\mathbb{R}^{d \times \dots \times d}$. We can interpret a k -tensor on \mathbb{R}^d as a multilinear map into \mathbb{R} with k arguments from \mathbb{R}^d . Hence, for k -tensor T on \mathbb{R}^d (with $k \geq 2$) and a vector $v \in \mathbb{R}^d$, we define $Tv := T(v, \dots)$ as a $(k-1)$ -tensor on \mathbb{R}^d . Using this, we can iteratively define the operator norm of T as

$$\|T\|_{\text{op}} := \begin{cases} \|T\|_2, & \text{if } T \in \mathbb{R}^d, \\ \sup_{\substack{v \in \mathbb{R}^d, \\ \|v\|_2 \leq 1}} \|Tv\|_{\text{op}}, & \text{if } T \text{ is a } k\text{-tensor on } \mathbb{R}^d \text{ with } k \geq 2. \end{cases}$$

With a slight abuse of notation, we will write $\|\cdot\|_2$ instead of $\|\cdot\|_{\text{op}}$ for all k -tensors.

2.1 Assumptions on f

The potential function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is assumed to be differentiable throughout this paper. In addition, we will always assume that f to be m -strongly convex

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}m\|x - y\|_2^2, \quad (\text{A.1})$$

and that the gradient $\nabla f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is M_1 -Lipschitz continuous

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq M_1\|x - y\|_2, \quad (\text{A.2})$$

for all $x, y \in \mathbb{R}^d$. It is straightforward to show that the two conditions (A.1) and (A.2) are equivalent to the Hessian of f being positive definite and satisfying $mI_d \preceq \nabla^2 f(x) \preceq M_1I_d$ (where I_d is the $d \times d$ identity matrix and $A \preceq B$ means that $x^T A x \leq x^T B x$ for $x \in \mathbb{R}^d$).

To establish a second order rate of convergence for the QUICSORT method, we shall assume further that f is twice differentiable and that its Hessian $\nabla^2 f$ is M_2 -Lipschitz continuous

$$\|\nabla^2 f(x)v - \nabla^2 f(y)v\|_2 \leq M_2\|x - y\|_2 \|v\|_2, \quad (\text{A.3})$$

and, to establish the third order convergence in Theorem 2, we additionally assume that f is three times differentiable and that its third derivative $\nabla^3 f$ is M_3 -Lipschitz continuous

$$\|\nabla^3 f(x)(v, u) - \nabla^3 f(y)(v, u)\|_2 \leq M_3\|x - y\|_2 \|v\|_2 \|u\|_2, \quad (\text{A.4})$$

for all $x, y, u, v \in \mathbb{R}^d$.

2.2 Probability notation

Suppose $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}_{t \geq 0})$ is a filtered probability space carrying a standard d -dimensional Brownian motion. The only SDE that we study in this paper is (2) which, under our smoothness and convexity assumptions on f , admits a unique strong solution that is ergodic and whose stationary measure π has a density $\pi(x, v) \propto e^{-f(x) - \frac{1}{2u}\|v\|_2^2}$. The numerical methods for (2) are computed at times $\{t_n\}_{n \geq 0}$ with $t_0 = 0$, $t_{n+1} > t_n$ and $h_n = t_{n+1} - t_n$. However, for deriving our global error estimates, we assume the step sizes $\{h_n\}$ are constant. For a random variable X , taking values in either \mathbb{R}^d or \mathbb{R}^{2d} , we define the \mathbb{L}_p norm of X as

$$\|X\|_{\mathbb{L}_p} := \mathbb{E}[\|X\|_2^p]^{\frac{1}{p}},$$

for all $p \geq 1$.

2.3 Coupling and Wasserstein distance

Let μ, ν be probability measures on \mathbb{R}^d . A coupling between μ and ν is a random variable $Z = (X, Y)$ for which $X \sim \mu$ and $Y \sim \nu$. For $p \geq 1$, we define the p -Wasserstein distance between μ and ν as

$$W_p(\mu, \nu) := \inf_{(X, Y) \sim (\mu \times \nu)} \|X - Y\|_{\mathbb{L}_p},$$

where the above infimum is taken over all couplings of the random variables X and Y with distributions $X \sim \mu$ and $Y \sim \nu$.

In this paper, we are interested in obtaining non-asymptotic bounds for the 2-Wasserstein distance between the target distribution $p(x) \propto e^{-f(x)}$ and the QUICSORT method (3). For this, we will use the standard bound between the distribution of X_n and $p(x) \propto e^{-f(x)}$,

$$W_2(\text{Law}(X_n), p) \leq \|X_n - x_{t_n}\|_{\mathbb{L}_2}, \quad (5)$$

where (x_t, v_t) denotes the solution of the SDE (2) started from the stationary distribution $(x_0, v_0) \sim \pi$. This then implies that $x_t \sim p$ for all $t \geq 0$, which is required for (5) to hold. Both the SDE solution (x_t, v_t) and (X_n, V_n) are obtained using the same Brownian motion.

2.4 Definition of the variables W_n , H_n and K_n

When Taylor expanding the SDE (2), we will obtain iterated integrals of Brownian motion. So, in order to derive a high order scheme, we require a method of simulating these integrals. In Foster et al. (2021), the authors use random variables (W_n, H_n, K_n) , which we now define.

Definition 3 Let $\{t_n\}_{n \geq 0}$ be a sequence of times with $t_0 = 0$, $t_{n+1} > t_n$ and $h_n = t_{n+1} - t_n$. From the same Brownian motion $W = \{W_t\}$, we define the following three random variables,

$$\begin{aligned} W_n &:= W_{t_n, t_{n+1}} = W_{t_{n+1}} - W_{t_n}, \\ H_n &:= \frac{1}{h_n} \int_{t_n}^{t_{n+1}} \left(W_{t_n, t} - \frac{t - t_n}{h_n} W_n \right) dt, \\ K_n &:= \frac{1}{h_n^2} \int_{t_n}^{t_{n+1}} \left(\frac{1}{2} h_n - (t - t_n) \right) \left(W_{t_n, t} - \frac{t - t_n}{h_n} W_n \right) dt. \end{aligned} \quad (6)$$

Theorem 4 Let I_d denote the $d \times d$ identity matrix. Then the three random variables, $W_n \sim \mathcal{N}(0, h_n I_d)$, $H_n \sim \mathcal{N}(0, \frac{1}{12} h_n I_d)$ and $K_n \sim \mathcal{N}(0, \frac{1}{720} h_n I_d)$ are independent and

$$\begin{aligned} \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 &= \frac{1}{2} h_n W_n + h_n H_n, \\ \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 &= \frac{1}{6} h_n^2 W_n + \frac{1}{2} h_n^2 H_n + h_n^2 K_n. \end{aligned} \quad (7)$$

Proof The result was shown in Foster (2024, Theorem 2 and Remark 6). ■

For a detailed derivation of these variables, we refer the reader to Foster et al. (2020) and Foster (2024, Section 2). The key idea underlying our QUICSORT method and its third order convergence is the same idea underlying the SORT method in Foster et al. (2021). Namely, we will replace the Brownian motion driving the SDE with a piecewise linear path. By constructing the piecewise linear path \mathcal{W} on $[t_n, t_{n+1}]$ to match the increment W_n and the integrals (7), the resulting ODE approximation matches the desired terms in its Taylor expansion and achieves high order convergence. This was shown in Foster et al. (2021). More generally, this approach can be used to design high order splitting methods for SDEs (see Foster et al. (2024)). In our case, we will define the following piecewise linear path \mathcal{W} .

Definition 5 Let $\{t_n\}_{n \geq 0}$ be a sequence of times with $t_0 = 0$, $t_{n+1} > t_n$ and $h_n = t_{n+1} - t_n$. We define an \mathbb{R}^d -valued piecewise linear path $\mathcal{W} = \{\mathcal{W}_t\}_{t \geq 0}$, whose n -th piece is given by

$$\mathcal{W}_t := \begin{cases} (H_n + 6K_n) + (W_n - 12K_n) \frac{t-t_n}{h_n}, & \text{for } t \in (t_n, t_{n+1}) \\ W_t, & \text{for } t \in \{t_n, t_{n+1}\}. \end{cases}$$

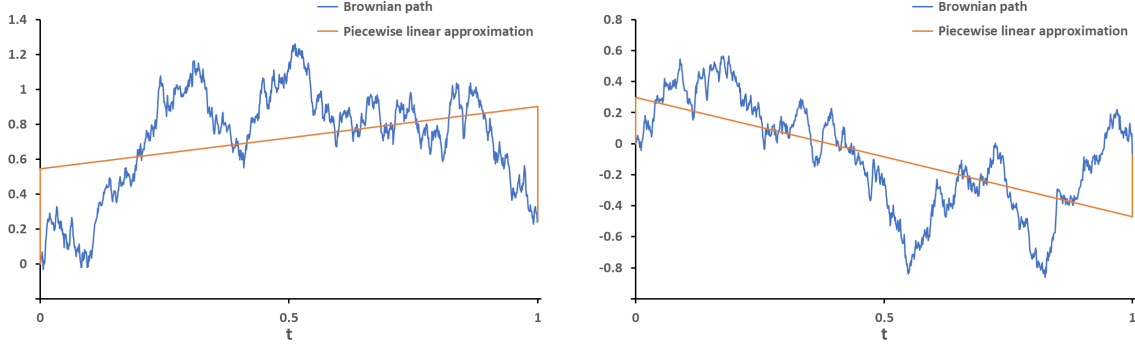


Figure 2: Brownian motion alongside the piecewise linear path \mathcal{W} given by Definition 5 (diagram taken from Foster et al. (2021)).

2.5 Derivation of the QUICSORT method

In Foster et al. (2021), the authors construct a splitting method for the underdamped Langevin dynamics (2) using the piecewise linear path \mathcal{W} defined in the previous subsection. We will recall this splitting method and then use it to derive the QUICSORT method (3).

Definition 6 (The shifted ODE approximation of ULD) Let $\{t_n\}_{n \geq 0}$ be a sequence of times with $t_0 = 0$, $t_{n+1} > t_n$ and $h_n = t_{n+1} - t_n$. We construct a numerical solution $\{(X_n, V_n)\}_{n \geq 0}$ by first setting some initial $(X_0, V_0) \sim \pi_0$ and for each $n \geq 0$ defining (X_{n+1}, V_{n+1}) as

$$\begin{pmatrix} X_{n+1} \\ V_{n+1} \end{pmatrix} := \begin{pmatrix} \tilde{x}_{t_{n+1}}^n \\ \tilde{v}_{t_{n+1}}^n \end{pmatrix} - (H_n - 6K_n) \begin{pmatrix} 0 \\ \sigma \end{pmatrix},$$

where $\sigma := \sqrt{2\gamma u}$ and $\{(\tilde{x}_t^n, \tilde{v}_t^n)\}_{t \in [t_n, t_{n+1}]}$ solves the following Langevin-type ODE

$$\frac{d}{dt} \begin{pmatrix} \tilde{x}_t^n \\ \tilde{v}_t^n \end{pmatrix} = \begin{pmatrix} \tilde{v}_t^n \\ -\gamma \tilde{v}_t^n - u \nabla f(\tilde{x}_t^n) \end{pmatrix} + (W_n - 12K_n) \begin{pmatrix} 0 \\ \sigma \end{pmatrix}, \quad (8)$$

with initial condition

$$\begin{pmatrix} \tilde{x}_{t_n}^n \\ \tilde{v}_{t_n}^n \end{pmatrix} := \begin{pmatrix} X_n \\ V_n \end{pmatrix} + (H_n + 6K_n) \begin{pmatrix} 0 \\ \sigma \end{pmatrix}.$$

We now wish to find a suitable numerical discretisation of the Langevin-type ODE (8). We first let $C_n := \sigma(W_n - 12K_n)$ and define a change of variables $\tilde{w}_t^n := e^{\gamma t} \tilde{v}_t^n$. This gives,

$$\frac{d}{dt} \begin{pmatrix} \tilde{x}_t^n \\ \tilde{w}_t^n \end{pmatrix} = \begin{pmatrix} e^{-\gamma t} \tilde{w}_t^n \\ -u e^{\gamma t} \nabla f(\tilde{x}_t^n) + \sigma e^{\gamma t} C_n \end{pmatrix}.$$

Taking the integral of both sides and expressing everything in terms of \tilde{x}^n and \tilde{v}^n produces

$$\begin{aligned}\tilde{x}_{t_{n+1}}^n &= \tilde{x}_{t_n}^n + \frac{1 - e^{-\gamma h_n}}{\gamma} \tilde{v}_{t_n}^n - u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} e^{-\gamma(r_1-r_2)} \nabla f(\tilde{x}_{r_2}^n) dr_2 dr_1 + \sigma \frac{e^{-\gamma h_n} + \gamma h_n - 1}{\gamma^2 h_n} C_n, \\ \tilde{v}_{t_{n+1}}^n &= e^{-\gamma h_n} \tilde{v}_{t_n}^n - u \int_{t_n}^{t_{n+1}} e^{-\gamma(t_{n+1}-r_1)} \nabla f(\tilde{x}_{r_1}^n) dr_1 + \sigma \frac{1 - e^{-\gamma h_n}}{\gamma h_n} C_n.\end{aligned}$$

We then approximate the integral terms using a two-point Gauss–Legendre quadrature. That is, we use the quadrature points $\lambda_{\pm} = \frac{3 \pm \sqrt{3}}{6}$ and obtain

$$\int_{t_n}^{t_{n+1}} e^{-\gamma(t_{n+1}-r_1)} \nabla f(\tilde{x}_{r_1}^n) dr_1 \approx \frac{1}{2} h_n \left(e^{-\lambda_+ \gamma h_n} \nabla f(\tilde{x}_{t_n + \lambda_- h_n}^n) + e^{-\lambda_- \gamma h_n} \nabla f(\tilde{x}_{t_n + \lambda_+ h_n}^n) \right),$$

and by additionally applying Fubini's theorem

$$\begin{aligned}& \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} e^{-\gamma(r_1-r_2)} \nabla f(\tilde{x}_{r_2}^n) dr_2 dr_1 \\ &= \int_{t_n}^{t_{n+1}} \int_{r_2}^{t_{n+1}} e^{-\gamma(r_1-r_2)} dr_1 \nabla f(\tilde{x}_{r_2}^n) dr_2 \\ &\approx \frac{1}{2} h_n \left(\frac{1 - e^{-\lambda_+ \gamma h_n}}{\gamma} \nabla f(\tilde{x}_{t_n + \lambda_- h_n}^n) + \frac{1 - e^{-\lambda_- \gamma h_n}}{\gamma} \nabla f(\tilde{x}_{t_n + \lambda_+ h_n}^n) \right).\end{aligned}$$

The terms $X_n^{(1)}$ and $X_n^{(2)}$ in the QUICSORT method are then designed to approximate $\tilde{x}_{t_n + \lambda_-}^n$ and $\tilde{x}_{t_n + \lambda_+}^n$. To reduce the number of gradient evaluations, the formula for $X^{(1)}$ does not involve ∇f . This is inspired by the UBU scheme (Sanz-Serna and Zygalkakis, 2021) which approximates the SDE solution at the midpoint of the interval before evaluating ∇f . But, as $X^{(1)}$ has no ∇f term, the resulting scheme does not match the desired expansion. Thus, to correct this, we use a coefficient of $\frac{1}{3}$ instead of λ_+ for the gradient term in $X^{(2)}$.

As mentioned above, the QUICSORT method is a natural extension to the UBU scheme, with the difference being that we use a two-point Gaussian quadrature to approximate the gradient terms instead of a midpoint approximation. Similarly, just as the UBU scheme has a contractivity property for strongly convex f , we can show QUICSORT is also contractive. This is precisely the reason why we can establish third order convergence guarantees for the QUICSORT method and not the SORT method originally proposed in Foster et al. (2021).

3. Error analysis of the QUICSORT method

In this section, we will present our main results regarding the 2-Wasserstein convergence of our new method (3) to the stationary distribution of (2). Our main theorem is given below.

Theorem 7 (Global error bound) *Let $\{(X_n, V_n)\}_{n \geq 0}$ be the QUICSORT method (3) and let $\{(x_t, v_t)\}_{t \geq 0}$ denote the underdamped Langevin diffusion, given by (2), with the same underlying Brownian motion. Suppose that the assumptions (A.1) and (A.2) hold. Suppose further that $(x_0, v_0) \sim \pi$, the unique stationary measure of the diffusion process*

and that both processes have the same initial velocity $V_0 = v_0 \sim \mathcal{N}(0, uI_d)$. Let $0 < h \leq 1$ be fixed and $\gamma \geq 2\sqrt{uM_1}$. Then for all $n \geq 0$, the global error at time $t = t_n$ has the bound

$$\|X_n - x_{t_n}\|_{\mathbb{L}_2} \leq ce^{-n\alpha h}\|X_0 - x_0\|_{\mathbb{L}_2} + C_1\sqrt{d}h, \quad (9)$$

where the contraction rate α is defined as

$$\alpha = \frac{1}{\gamma} \min \left(\gamma^2 - uM_1, um, \lambda_- \gamma^2 - \frac{1}{2}u(M_1 - m) \right) - Kh, \quad (10)$$

and the constants $c, C_1, K > 0$ depend only on γ, u, m and M_1 . Therefore, if h is sufficiently small, we have $\alpha > 0$. Under the additional assumption (A.3), the error at time t_n satisfies

$$\|X_n - x_{t_n}\|_{\mathbb{L}_2} \leq ce^{-n\alpha h}\|X_0 - x_0\|_{\mathbb{L}_2} + C_2dh^2, \quad (11)$$

where $C_2 > 0$ is a constant depending on γ, u, m, M_1 and M_2 .

Finally, under the additional assumption (A.4), the global error at time $t = t_n$ has the bound

$$\|X_n - x_{t_n}\|_{\mathbb{L}_2} \leq ce^{-n\beta h}\|X_0 - x_0\|_{\mathbb{L}_2} + C_3d^{1.5}h^3, \quad (12)$$

where $\beta = \frac{1}{2}\alpha$ and $C_3 > 0$ is a constant depending on γ, u, m, M_1, M_2 and M_3 .

A proof for this theorem can be found in Appendix D. The various constants are given explicitly throughout the appendices. The general recipe for the error analysis is as follows.

- At each time step t_{n+1} , we wish to bound the \mathbb{L}_2 error between the underdamped Langevin diffusion (2) and the QUICSORT method (3) where we assume the two processes are synchronously coupled (i.e. driven by the same Brownian motion W).
- We introduce a third process, obtained by applying the QUICSORT approximation with step size h_n to the diffusion process at each time t_n .
- By showing that the QUICSORT method is contractive, we obtain an exponentially decaying bound for the \mathbb{L}_2 distance between our approximation and this third process.
- We establish local error bounds between the true diffusion (2) and the third process. This allows us to estimate the \mathbb{L}_2 error at time t_{n+1} between the true diffusion (2) and the QUICSORT approximation (3) in terms of the same error at time t_n .
- By iteratively unfolding this local error bound, we obtain a global \mathbb{L}_2 error estimate over the entire length of the Markov chain. In particular, due to the contractivity of the QUICSORT method when f is strongly convex, this bound holds for all $n \geq 0$.

This approach is illustrated in Figure 3:

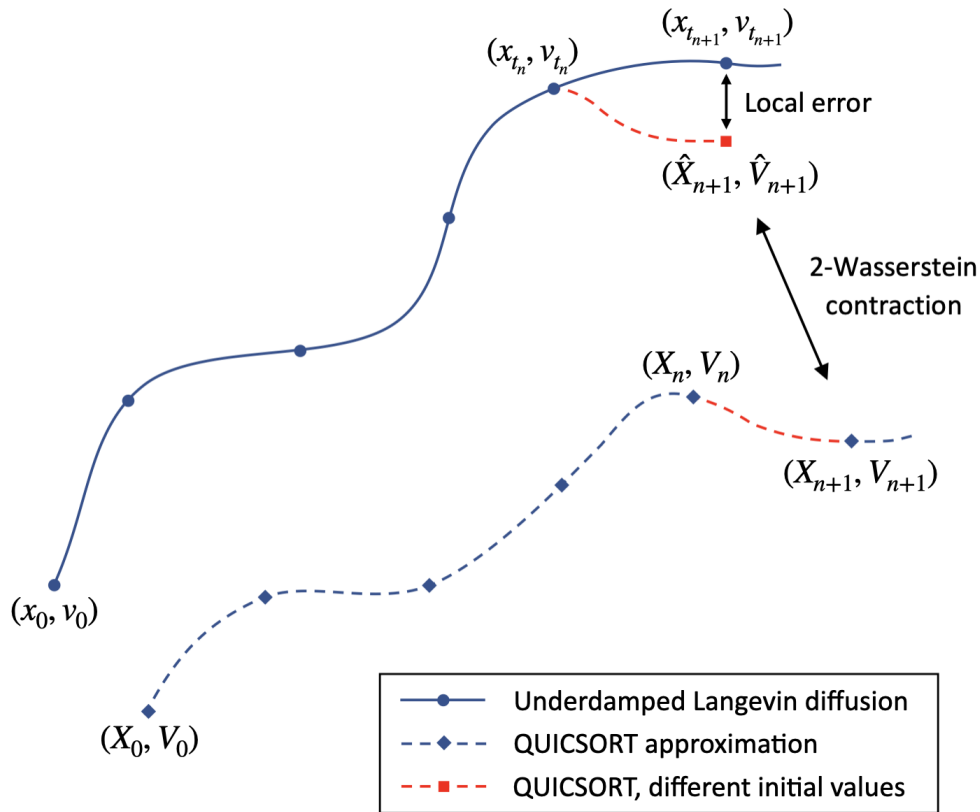


Figure 3: Diagram outlining the general strategy for our error analysis.

For the rest of the section, we outline the key results used in each step of the error analysis.

3.1 Exponential contractivity

Typically the local errors between the approximation and the diffusion process propagate like a random walk and we would not be able to bound the global \mathbb{L}_2 error for all $n \geq 0$. Fortunately, it is possible to remedy this by applying a suitable \mathbb{L}_2 contractivity argument. For ULD, contractivity does not hold directly for the solution process $\{(x_t, v_t)\}_{t \geq 0}$, but instead holds after introducing a coordinate transformation (Eberle et al., 2019). In this paper we use the approach introduced by Dalalyan and Riou-Durand (2020) which considers,

$$\left\{ \begin{pmatrix} wx_t + v_t \\ zx_t + v_t \end{pmatrix} \right\}_{t \geq 0},$$

where $w \in [0, \frac{1}{2}\gamma)$ and $z := \gamma - w$ to obtain \mathbb{L}_2 contractivity. However, to establish our higher order convergence guarantees, we prove a similar contraction result between two synchronously coupled steps of our QUICSORT method (3). This allows us to bound the local error at each time step in terms of global \mathbb{L}_p bounds of the underlying diffusion process.

Our main result for this step is as follows.

Theorem 8 (Exponential contractivity of the QUICSORT method) *Consider the application of the QUICSORT method from different initial conditions (X_n, V_n) and (Y_n, U_n) at time t_n . Let (X_{n+1}, V_{n+1}) and (Y_{n+1}, U_{n+1}) be the corresponding approximations obtained at time t_{n+1} using the same Brownian motion. We define the coordinate transformations:*

$$\begin{aligned} \begin{pmatrix} W_n \\ Z_n \end{pmatrix} &= \begin{pmatrix} (wX_n + V_n) - (wY_n + U_n) \\ (zX_n + V_n) - (zY_n + U_n) \end{pmatrix}, \\ \begin{pmatrix} W_{n+1} \\ Z_{n+1} \end{pmatrix} &= \begin{pmatrix} (wX_{n+1} + V_{n+1}) - (wY_{n+1} + U_{n+1}) \\ (zX_{n+1} + V_{n+1}) - (zY_{n+1} + U_{n+1}) \end{pmatrix}, \end{aligned} \quad (13)$$

where $w \in [0, \frac{1}{2}\gamma)$ and $z = \gamma - w$. Then under the assumptions (A.1) and (A.2), we have

$$(\|W_{n+1}\|_{\mathbb{L}_2} + \|Z_{n+1}\|_{\mathbb{L}_2}) \leq e^{-\alpha h_n} (\|W_n\|_{\mathbb{L}_2} + \|Z_n\|_{\mathbb{L}_2}), \quad (14)$$

for all $n \geq 0$ and $h_n \in (0, 1]$, where the contraction rate α is given by

$$\alpha = \frac{1}{\gamma - 2w} \min \left((\gamma - w)^2 - uM_1, um - w^2, (\lambda_-(\gamma - w)^2 - \lambda_+w^2) - \frac{1}{2}u(M_1 - m) \right) - Kh_n, \quad (15)$$

for some constant $K > 0$ not depending on h_n .

A proof of the above can be found in Appendix B.

3.2 Global \mathbb{L}_p bounds for underdamped Langevin dynamics

Over each time interval $[t_n, t_{n+1}]$, we wish to iteratively bound the error at t_{n+1} in terms of the error at time t_n . In order to obtain estimates independently of the current position and momentum of the SDE, we use the global \mathbb{L}_p bounds of underdamped Langevin dynamics. In particular, by using the moments of the momentum's stationary distribution ($\propto e^{-f}$) and extending Dalalyan (2017, Lemma 2), we will establish the following global \mathbb{L}_p bounds.

Theorem 9 (Global \mathbb{L}_p bounds for ULD) *Let $\{(x_t, v_t)\}_{t \geq 0}$ be the underdamped Langevin diffusion (2) with initial condition $(x_0, v_0) \sim \pi$, the stationary distribution of the process. Then, for all $p \geq 1$,*

$$\|v_t\|_{\mathbb{L}_{2p}} \leq C(p)\sqrt{ud}, \quad (16)$$

where $C(p) = [(2p - 1)!!]^{\frac{1}{2p}}$. Here, $!!$ denotes the standard double factorial for odd numbers. In addition, under the assumption that the gradient of f is M_1 -Lipschitz continuous, we have

$$\|\nabla f(x_t)\|_{\mathbb{L}_{2p}} \leq C(p)\sqrt{M_1 d}, \quad (17)$$

for all $p \geq 1$.

A proof of the above can be found in Appendix A.

3.3 Local error estimates

We wish to estimate the absolute average difference (local weak error) and the mean squared difference (local strong error) between ULD and its QUICSORT approximation over the interval $[t_n, t_{n+1}]$ when the processes are synchronously coupled and coincide at time t_n . By applying a contraction argument to the approximation and not to the SDE itself, we can use the global \mathbb{L}_p bounds for the diffusion process when bounding each local error. Furthermore, under the smoothness assumptions on the scalar potential f , we are also able to estimate all of the error terms involving derivatives of f .

Theorem 10 (Local error bounds) *Let $Q_n = (X_n, V_n)$ be the QUICSORT scheme (3) and $q_t = (x_t, v_t)$ denote the underdamped Langevin diffusion (2) with the same underlying Brownian motion. Suppose the assumptions (A.1) and (A.2) hold and assume further that $(x_0, v_0) \sim \pi$, the unique stationary distribution of the diffusion process. Let $0 < h_n \leq 1$ be fixed. Then, if $Q_n = q_{t_n}$ at time t_n , the local strong and weak errors at time t_{n+1} are bounded by*

$$\|Q_{n+1} - q_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_s^{(1)} \sqrt{d} h_n^2, \quad \|\mathbb{E}[Q_{n+1} - q_{t_{n+1}}]\|_2 \leq C_w^{(1)} \sqrt{d} h_n^2, \quad (18)$$

where $C_s^{(1)}, C_w^{(1)} > 0$ are finite constants depending on γ, u, m and M_1 .

Under the additional assumption (A.3), the local strong and weak errors are bounded by

$$\|Q_{n+1} - q_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_s^{(2)} d h_n^3, \quad \|\mathbb{E}[Q_{n+1} - q_{t_{n+1}}]\|_2 \leq C_w^{(2)} d h_n^3, \quad (19)$$

where $C_s^{(2)}, C_w^{(2)} > 0$ are finite constants depending on γ, u, m, M_1 and M_2 .

Finally, under the additional assumption (A.4), the local errors are bounded by

$$\|Q_{n+1} - q_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_s^{(3)} d^{1.5} h_n^{3.5}, \quad \|\mathbb{E}[Q_{n+1} - q_{t_{n+1}}]\|_2 \leq C_w^{(3)} d^{1.5} h_n^4, \quad (20)$$

where $C_s^{(3)}, C_w^{(3)} > 0$ are finite constants depending on γ, u, m, M_1, M_2 and M_3 .

A proof for the above theorem can be found in Appendix C.

4. Numerical Experiments

In this section, we will present experiments to demonstrate the convergence properties of the proposed QUICSORT method (3). Firstly, we shall compare the QUICSORT method to other approximations of ULD and empirically verify its third order convergence rate. Secondly, we will show the efficacy of QUICSORT as an MCMC algorithm for performing Bayesian logistic regression on real-world data. In particular, in our highest dimensional datasets, we observe that QUICSORT (and the UBU splitting) have faster mixing times than the popular “No U-Turn Sampler” (NUTS) introduced in Hoffman and Gelman (2014).

The code for our experiments can be found at github.com/andyElking/ThirdOrderLMC. The ULD-based MCMC sampling was done by simulating ULD using DiffraX (Kidger, 2021) and we used the package NumPyro (Phan et al., 2019) to generate samples using NUTS.

To begin, we will define the logistic regression model that we aim to generate samples from. Given a labelled dataset $\mathcal{D} = \{(x_i, y_i)\}_{1 \leq i \leq m}$ with features $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$, the logistic regression aims to find weights $\theta \in \mathbb{R}^d$ and a bias $b \in \mathbb{R}$ such that the relation

$$\mathbb{P}_{\theta, b}(y_i = 1 | x_i) = \frac{1}{1 + \exp(-(\langle \theta, x_i \rangle + b))},$$

is a good model for the data. Specifically, we would like to maximise the likelihood

$$p(\mathcal{D} | \theta, b) = \prod_{i=1}^m \frac{1}{1 + \exp(-y_i(\langle \theta, x_i \rangle + b))}. \quad (21)$$

For Bayesian logistic regression, we impose a prior distribution on the parameters $\tilde{\theta} := (\theta, b)$ and aim to generate samples from the posterior distribution $p(\tilde{\theta} | \mathcal{D})$ given by Bayes’ rule:

$$p(\tilde{\theta} | \mathcal{D}) = \frac{p(\mathcal{D} | \tilde{\theta})p(\tilde{\theta})}{p(\mathcal{D})}. \quad (22)$$

In this section, we will use the following Gaussian prior distribution for the parameters $\tilde{\theta}$,

$$\tilde{\theta} \sim p(\tilde{\theta}) \iff \theta \sim \mathcal{N}\left(0, \frac{1}{2\text{Var}(\{x_i\})} I_d\right), \quad b \sim \mathcal{N}(0, 1), \quad (23)$$

where $\text{Var}(\{x_i\})$ is the empirical variance of the dataset $\{x_1, \dots, x_m\}$. We rescale by this variance simply as a way to normalise the inner products in (21). Whilst computing $p(\mathcal{D})$ involves an intractable integral, it disappears when we compute $\nabla f(\tilde{\theta}) = \nabla(-\log(p(\tilde{\theta} | \mathcal{D})))$.

Therefore, using Bayes’ rule (22), the scalar potential coming from our logistic regression is

$$f(\tilde{\theta}) := \sum_{i=1}^m \log(1 + \exp(-y_i(\langle \theta, x_i \rangle + b))) + \frac{1}{4\text{Var}(\{x_i\})} \|\theta\|_2^2 + \frac{1}{2} b^2. \quad (24)$$

Remark 11 *The logistic regression loss function is typically convex but not strongly convex. However, due to the quadratic regularisation terms appearing due to the prior distribution, the potential (24) is strongly convex. In addition, just as for the standard logistic regression loss, the above potential is smooth and its derivatives are all globally Lipschitz continuous. Thus, it satisfies the assumptions needed in our error analysis of QUICSORT (Theorem 7).*

4.1 Verifying the third order convergence of QUICSORT

In this subsection, we will present our experiment for empirically verifying the convergence rates for some of the Langevin methods given in Table 1 (including the QUICSORT method).

We will be simulating the following underdamped Langevin dynamics with $\gamma = 1$ and $u = 1$,

$$\begin{aligned} d\tilde{\theta}_t &= v_t dt, \\ dv_t &= -v_t dt - \nabla f(\tilde{\theta}_t) dt + \sqrt{2} dW_t, \end{aligned} \tag{25}$$

where the initial condition $\tilde{\theta}_0$ is sampled from the prior distribution (23) and $v_0 \sim \mathcal{N}(0, I_d)$. The scalar potential f comes from the previously described Bayesian logistic regression applied to the Titanic passenger survival dataset (Hendricks, 2015) and is of the form (24). The dimensionality of the dataset is 4, so the corresponding Langevin dynamics has $\tilde{\theta}_t \in \mathbb{R}^5$.

To estimate the strong (or \mathbb{L}_2) convergence rate of each numerical method, we first compute reference solutions by simulating the SDE (25) using a very fine step size (in our experiment, we use $h_{\text{fine}} = 2^{-16}T$). More precisely, since we are estimating the \mathbb{L}_2 error by Monte Carlo simulation, we will compute a reference solution for each sample path of Brownian motion. We used QUICSORT to compute these reference solutions due to its fast convergence rate.

For each Brownian sample path, we then simulate (25) using each numerical method with a larger step size and compute the mean squared error against the corresponding reference solution at time T . By averaging over sample paths, we obtain the following estimator for $\mathbb{E}[\|X_N - \tilde{\theta}_T\|_2^2]$ where X_N is the numerical approximation for the SDE solution at time T .

Definition 12 (Strong error estimator) *Let $T > 0$ be fixed. For $N \geq 1$, we compute a numerical solution $\{(X_n, V_n)\}_{0 \leq n \leq N}$ of (25) at times $t_n := nh$ using a step size of $h = \frac{T}{N}$. Similarly, let $\{(X_n^{\text{fine}}, V_n^{\text{fine}})\}_{n \geq 0}$ be an approximation obtained with a smaller step size h_{fine} . We generate J samples of $\{(X, V, X^{\text{fine}}, V^{\text{fine}})\}$ and compute the Monte Carlo estimator:*

$$S_{N,J} := \sqrt{\frac{1}{J} \sum_{j=1}^J \|X_{N,j} - X_{N,j}^{\text{fine}}\|_2^2}, \tag{26}$$

where each pair $(X_{N,j}, X_{N,j}^{\text{fine}})$ is computed using the same sample path of Brownian motion and the initial condition $X_{0,j} = X_{0,j}^{\text{fine}}$ is sampled from the Gaussian prior distribution (23).

Remark 13 *To simulate the SDE (25), we use the DiffraX software package (Kidger, 2021). Given a single PRNG seed, the “Virtual Brownian Tree” in DiffraX can determine an entire sample path of Brownian motion (including the integrals (3)) up to a user-specified tolerance. This then makes it straightforward to generate multiple different numerical solutions using the same Brownian sample path. The various algorithmic details and Brownian bridge constructions used by DiffraX’s Virtual Brownian Tree can be found in Jelinčič et al. (2024).*

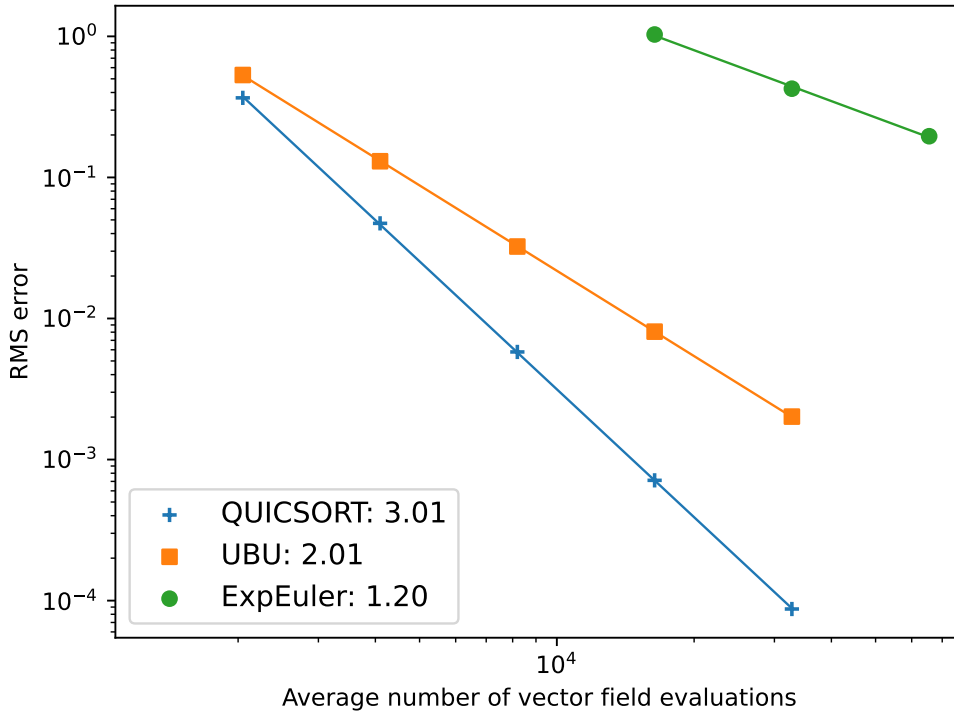


Figure 4: Empirical validation of the complexities in Table 1. Here, each numerical method is used to simulate the underdamped Langevin dynamics (25) where the potential f is obtained from a Bayesian logistic regression applied to the Titanic passenger survival dataset (Hendricks, 2015). The Monte Carlo estimator (26) for the root mean square (RMS) error is estimated at time $T = 100$ using $J = 1000$ samples. Instead of the number of steps, we report the number of evaluations of ∇f since that is often the most computationally expensive part of gradient-based MCMC.

From Figure 4, we can see that the numerical methods for ULD (25) achieve their theoretical rates of convergence with the QUICSORT method exhibiting a third order convergence rate.

When the step size is large, the QUICSORT method performs similarly to the UBU splitting method (Sanz-Serna and Zygalkis, 2021) which has only one gradient evaluation per step. However, as the step size becomes smaller, the third order convergence of QUICSORT makes it significantly more accurate than the UBU scheme – which is second order convergent. As one would expect, the exponential Euler scheme proposed in (Cheng et al., 2018) is the least accurate among the numerical methods and only shows a first order convergence rate.

However, estimating the strong convergence only gives an small indication of each methods’ efficacy as an MCMC algorithm. In the next subsection, we shall investigate this further.

4.2 Mixing time experiments against the No U-Turn Sampler (NUTS)

In this subsection, we compare the QUICSORT and UBU methods against the No U-Turn Sampler (NUTS) for performing Bayesian logistic regression with target distribution (22) and Gaussian prior (23). We test using 14 datasets with dimension d ranging from 3 to 617.

The first 13 datasets are those used in the Bayesian logistic experiments in Li et al. (2023). For the final dataset, we use the ISOLET dataset (Cole and Fanty, 1991) with $d = 617$ and 26 classes (letters of the alphabet), but we only use “A” and “B” for our logistic regression.

We would like to observe how quickly the MCMC chains approach the target distribution. Inspired by Li et al. (2023), we will compute the following metrics to assess this convergence.

- Energy distance

$$d^2(\mu, \nu) = 2\mathbb{E}_{X \sim \mu, Y \sim \nu}[\|X - Y\|_2] - \mathbb{E}_{X \sim \mu, X' \sim \mu}[\|X - X'\|_2] - \mathbb{E}_{Y \sim \nu, Y' \sim \nu}[\|Y - Y'\|_2],$$

where the random variables (X, Y, X', Y') within the expectations are independent. This is the Maximum Mean Discrepancy (MMD) obtained from the negative distance kernel $k(x, y) := -\|x - y\|_2$ and is explicitly computable when μ and ν are discrete.

- 2-Wasserstein distance

$$W_2(\mu, \nu) := \inf_{(X, Y) \sim (\mu \times \nu)} \|X - Y\|_{\mathbb{L}_2},$$

where the above infimum is taken over all couplings of the random variables X and Y with distributions $X \sim \mu$ and $Y \sim \nu$. If μ and ν are empirical distributions with J points each, it becomes very computationally expensive to find an optimal coupling between μ and ν (as J increases). For estimating the 2-Wasserstein distance between distributions, we use the Python Optimal Transport package (Flamary et al., 2021).

We present the results for 6 out of the 14 datasets in Figure 5. Here, we generate a collection of ‘ground truth’ samples using NUTS with a sufficiently large number of iterations and, at regular time intervals, compare against the samples from different MCMC algorithms (QUICSORT and UBU applied to ULD (25) as well as NUTS applied to the target (22)). Results for the other 8 datasets can be found at github.com/andyElking/ThirdOrderLMC.

In half of the datasets, we observed that NUTS converged significantly faster to the target distribution (22) than the numerical methods for ULD. This is perhaps unsurprising since NUTS is widely considered the “go-to” algorithm for performing gradient-based MCMC. However, in the two highest dimensional datasets, splice ($d = 61$) and isolet ($d = 617$), the Langevin numerical methods converge faster than NUTS (after an initial burn-in period). Hence, exploring these algorithms in higher dimensional settings is a topic of future work.

In these experiments, we were able to achieve good performance from the two Langevin numerical methods with a relatively large step size. However, as we observed in Figure 4, the strong accuracy of QUICSORT and UBU become very similar as the step size increases. This appears to carry over when we investigate the weak convergence of the algorithms as the QUICSORT and UBU methods give near identical performances across all 14 datasets. In these experiments, we set $h_{\text{QUICSORT}} = 2h_{\text{UBU}}$ so that the QUICSORT and UBU chains require the same number of gradient evaluations (i.e. to have essentially the same cost). Since both the energy and 2-Wasserstein distance become computationally expensive for large collections of particles, it is difficult to accurately compare the stationary distributions of the QUICSORT and UBU methods. Given the smoothness of the potential, we expect QUICSORT would have a better stationary distribution due to its faster convergence rate.

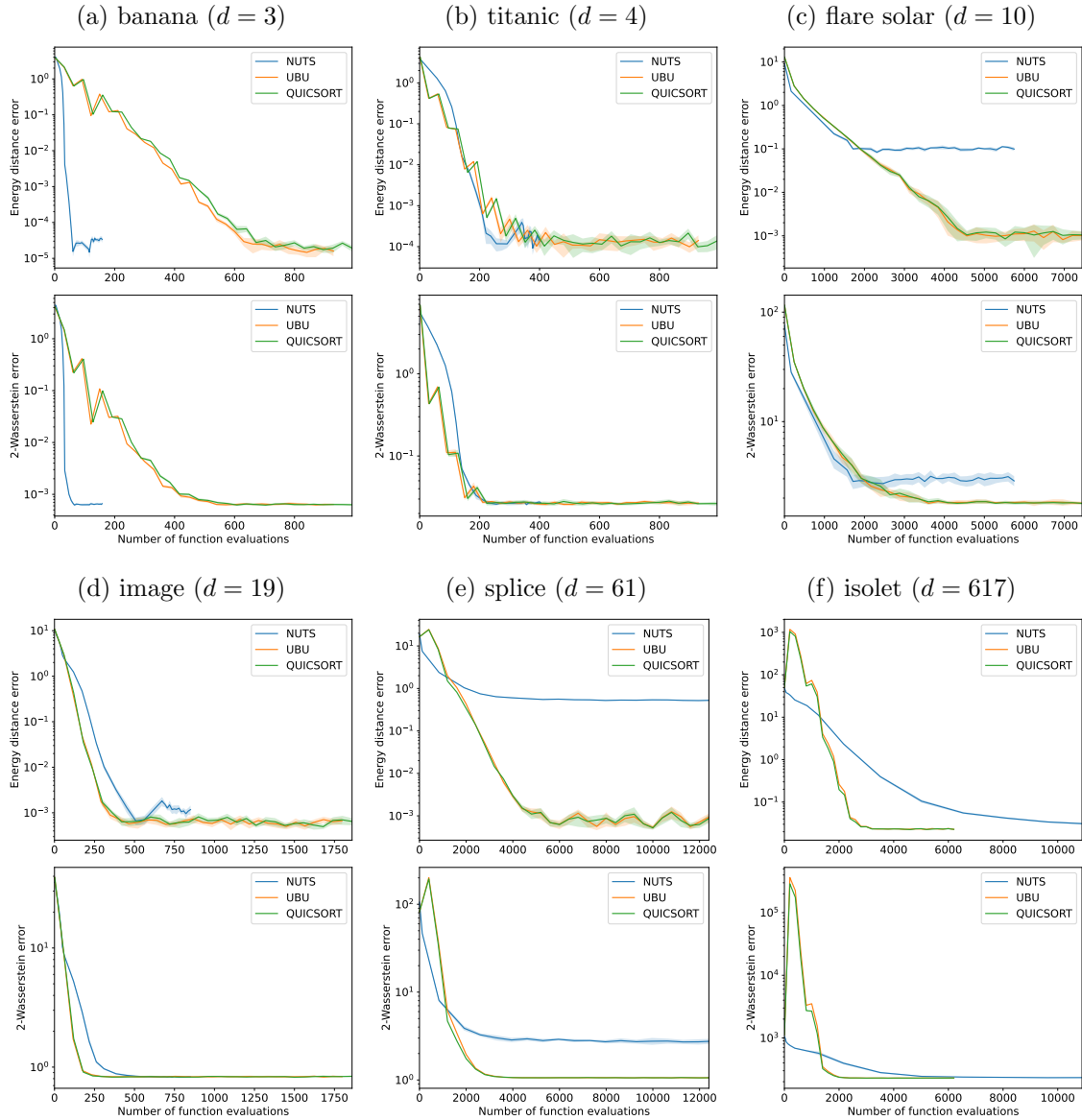


Figure 5: For each of the 14 datasets we ran 2^{15} independent chains of NUTS and the same number of Langevin trajectories simulated using the UBU and QUICSORT methods. At regular intervals along the run we compared these 2^{15} particles against the “ground truth” samples, which were obtained using NUTS with a sufficiently long burn-in period. We computed both the energy distance and the 2-Wasserstein distance (the latter on a subsample of size 2^{11}). These distances are plotted against the number of times the gradient ∇f was evaluated up to that point in the run. We present 6 of the 14 plots here and the rest can be found at github.com/andyElking/ThirdOrderLMC. We ran the experiment five times (with different PRNG seeds) and present the standard deviation as the shaded regions. Although different step sizes are used for different datasets, we choose the step size to be as large as possible and set $h_{\text{QUICSORT}} = 2h_{\text{UBU}}$ so that the ULD methods have similar cost.

5. Conclusion and future work

In this paper, we presented a numerical method for underdamped Langevin dynamics (ULD) called QUICSORT³, which can be seen as a particular discretisation of the “Shifted ODE” introduced in Foster et al. (2021). Under strong convexity and smoothness assumptions on the potential f , we obtain first, second and third order \mathbb{L}_2 error bounds for QUICSORT. In particular, by establishing a contraction property for QUICSORT, these non-asymptotic bounds hold over the entire trajectory of the numerical solutions. Therefore, when used as an MCMC algorithm to sample from a log-concave target distribution $p(x) \propto e^{-f(x)}$, we can bound the 2-Wasserstein error between p and the distribution of the QUICSORT samples.

Whilst similar bounds have been established for second order methods (such as for the UBU scheme by Sanz-Serna and Zygalkis (2021)), we believe that QUICSORT is the only ULD numerical method with third order convergence guarantees. Foster et al. (2021) obtained similar third order non-asymptotic bounds for the shifted ODE, but these bounds require the ODE to be solved exactly. Although other third order numerical methods for ULD have been proposed, they involve computing the Hessian $\nabla^2 f$ which is often challenging for high-dimensional applications. We refer the reader to Milstein and Tretyakov (2003); Shkerin and Sibiriyakov (2025) for examples of such derivative-based third order methods.

Finally, to support our theory, we presented numerical experiments where we demonstrated the third order convergence of QUICSORT and showed that it can outperform the widely-used No U-Turn Sampler (NUTS) as an MCMC algorithm for Bayesian logistic regression. However, in our MCMC experiments, we found that QUICSORT and UBU performed near identically across all 14 datasets. We expect this is simply due to the large step sizes used.

As well as considering more sophisticated Bayesian models, we would also like to explore the applications of Langevin-based MCMC algorithms with adaptive step sizes. In particular, Jelinčič et al. (2024) conduct additional Bayesian logistic regression experiments using the QUICSORT method with an adaptive step size controller. This was possible thanks to the “Virtual Brownian Tree” proposed by the authors and available in DiffraX (Kidger, 2021).

These adaptive step sizes are determined using Proportional Integral (PI) controllers, which are commonly used for numerical ODE simulation but have seen relatively little use for SDEs (Burrage et al., 2004; Ilie et al., 2015). Whilst QUICSORT with adaptive step sizes usually performed comparably to QUICSORT with constant step sizes, there were several datasets (including the two with the highest dimensions) where there was a significant improvement. These results are shown in Figure 6 and clearly demonstrates that PI step size controllers (which are very well supported by DiffraX) show great promise for Langevin-based MCMC.

In Chada et al. (2024), it is shown that the UBU method can be applied to perform unbiased MCMC estimation (even with inexact gradients) using Multilevel Monte Carlo techniques. Thus, a natural topic of future work would be to apply their methodology with QUICSORT. As a further topic, it may also be possible to apply their approach with adaptive step sizes.

Another direction of future research would be to apply the QUICSORT method to the Kinetic Interacting Particle Langevin Diffusion proposed by Valsecchi Oliva and Akyildiz (2024) for the joint estimation of latent variables and parameters in latent variable models.

3. QUadrature Inspired and Contractive Shifted ODE with Runge-Kutta Three

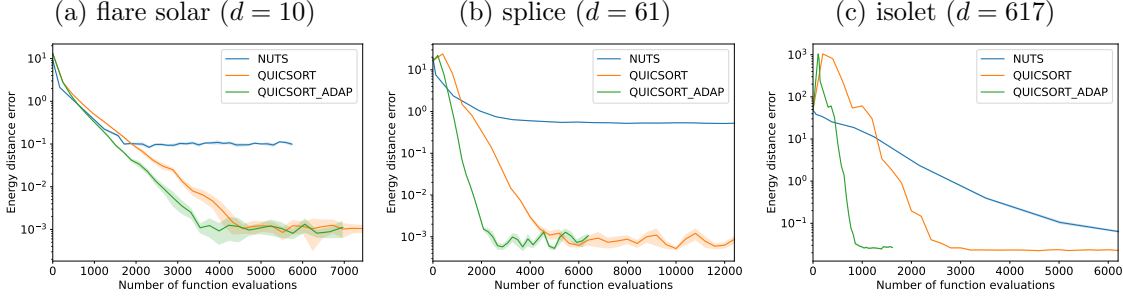


Figure 6: Using the same the target distribution and experimental setup as before, the QUICSORT method with constant and adaptive step sizes is now compared against NUTS. These adaptive step sizes are determined using a PI controller with parameters $K_P = 0.1$ and $K_I = 0.4$ (along with a user-specified tolerance and initial step size). We note that these are close to the parameters $(K_P, K_I) = (0.1, 0.3)$ recommended by Ilie et al. (2015). Similar to Jelinčič et al. (2024), we report the convergence results in the energy distance for three of the datasets (including the two highest dimensional ones), however the code and results for all 14 datasets can be found at github.com/andyElking/Single-seed_BrownianMotion.

Finally, another avenue of future research would be to improve the QUICSORT method by using “exponential” iterated integrals of the Brownian motion instead of (W_n, H_n, K_n) . This could then allow us to better match the following stochastic Taylor expansion of ULD:

$$\begin{aligned}
 x_{t_{n+1}} &= x_{t_n} + \left(\frac{1 - e^{-\gamma(t_{n+1}-t_n)}}{\gamma} \right) v_{t_n} + \int_{t_n}^{t_{n+1}} \int_s^{r_1} e^{-\gamma(r_1-r_2)} \nabla f(x_{r_2}) dr_2 dr_1 \\
 &\quad + \sqrt{2\gamma u} \int_{t_n}^{t_{n+1}} \int_s^{r_1} e^{-\gamma(r_1-r_2)} dW_{r_2} dr_1, \\
 v_{t_{n+1}} &= e^{-\gamma(t_{n+1}-t_n)} v_{t_n} + \int_{t_n}^{t_{n+1}} e^{-\gamma(t_{n+1}-r_1)} \nabla f(x_{r_1}) dr_1 + \sqrt{2\gamma u} \int_{t_n}^{t_{n+1}} e^{-\gamma(t_{n+1}-r_1)} dW_{r_1},
 \end{aligned}$$

In particular, we could construct a piecewise linear path \widetilde{W} on the interval $[s, t]$ such that

$$\begin{aligned}
 \int_s^t e^{-\gamma(t-r_1)} d\widetilde{W}_{r_1} &= \int_s^t e^{-\gamma(t-r_1)} dW_{r_1}, \\
 \int_s^t \int_s^{r_1} e^{-\gamma(r_1-r_2)} d\widetilde{W}_{r_2} dr_1 &= \int_s^t \int_s^{r_1} e^{-\gamma(r_1-r_2)} dW_{r_2} dr_1, \\
 \int_s^t e^{-\gamma(t-r_1)} \int_s^{r_1} \int_s^{r_2} e^{-\gamma(r_2-r_3)} d\widetilde{W}_{r_3} dr_2 dr_1 &= \int_s^t e^{-\gamma(t-r_1)} \int_s^{r_1} \int_s^{r_2} e^{-\gamma(r_2-r_3)} dW_{r_3} dr_2 dr_1.
 \end{aligned}$$

If the path $\{(r, \widetilde{W}_r)\}$ had increments $(0, A)$, $(t-s, B)$, $(0, C)$, then the first integral becomes

$$\int_s^t e^{-\gamma(t-r)} d\widetilde{W}_r = \underbrace{e^{-\gamma(t-s)} A}_{\text{vertical jump at } r=s} + \underbrace{\frac{1 - e^{-\gamma(t-s)}}{\gamma(t-s)} B}_{\text{linear piece}} + \underbrace{C}_{\text{jump at } r=t}.$$

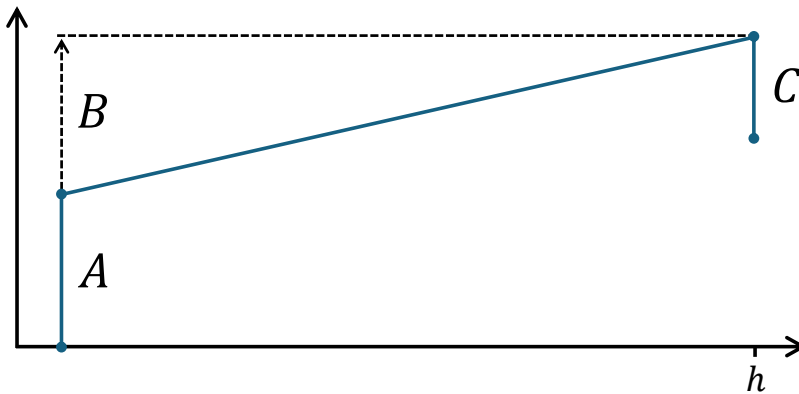


Figure 7: Illustration of the piecewise linear path \widetilde{W} with increments A , B and C to match “exponential” iterated integrals of Brownian motion. Here, $h = t - s$ denotes the step size.

Repeating for the other iterated integrals and substituting $h = t - s$, the problem of finding the increments A , B and C would reduce to solving the linear system $M\mathbf{x} = \mathbf{b}$ where

$$M = \begin{pmatrix} e^{-\gamma h} & \frac{1 - e^{-\gamma h}}{\gamma h} & 1 \\ \frac{1 - e^{-\gamma h}}{\gamma} & \frac{e^{-\gamma h} + \gamma h - 1}{\gamma^2 h} & 0 \\ \frac{1 - e^{-\gamma h}(1 + \gamma h)}{\gamma^2} & \frac{e^{-\gamma h}(2 + \gamma h) + \gamma h - 2}{\gamma^3 h} & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \int_s^t e^{-\gamma(t-r_1)} dW_{r_1} \\ \int_s^t \int_s^{r_1} e^{-\gamma(r_1-r_2)} dW_{r_2} dr_1 \\ \int_s^t e^{-\gamma(t-r_1)} \int_s^{r_1} \int_s^{r_2} e^{-\gamma(r_2-r_3)} dW_{r_3} dr_2 dr_1 \end{pmatrix}$$

and $\mathbf{x} = (A, B, C)^\top$. Since they are all linear functions of W , the iterated integrals in \mathbf{b} are jointly Gaussian, but their covariance matrix would need to be computed to generate them. Using these increments it is then straightforward to modify the QUICSORT method with $(H_n + 6K_n)$ replaced by A , $(W_n - 12K_n)$ replaced by B and $(H_n - 6K_n)$ replaced by C .

Acknowledgments

We greatly appreciate the support from the Department of Mathematical Sciences at the University of Bath, the Maths4DL programme under EPSRC grant EP/V026259/1 and the Alan Turing Institute. AJ thanks the UK National Cyber Security Centre for funding his research and Extropic for funding his work implementing high order SDE solvers in DiffraX.

References

- Alfonso Álamo and Jesus María Sanz-Serna. Word series for the numerical integration of stochastic differential equations: splitting integrators. *Communications on Pure and Applied Analysis*, 18(4):2163–2195, 2019.
- Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017.
- Houssam AlRachid, Letif Mones, and Christoph Ortner. Some remarks on preconditioning molecular dynamics. *SMAI Journal of Computational Mathematics*, 4:57–80, 2018.

- Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2013.
- Evelyn Buckwar, Massimiliano Tamborrino, and Irene Tubikanec. Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Statistics and Computing*, 30:627–648, 2019.
- Michael Bulmer. *Principles of Statistics*. Dover Publications, 1979.
- Pamela M. Burrage, Ratna Herdiana, and Kevin Burrage. Adaptive stepsize based on control theory for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 170(2):317–336, 2004.
- Giovanni Bussi and Michele Parrinello. Accurate sampling using Langevin dynamics. *Physical Review E*, 75(5):056707, 2006.
- Neil K. Chada, Benedict Leimkuhler, Daniel Paulin, and Peter A. Whalley. Unbiased Kinetic Langevin Monte Carlo with Inexact Gradients. *arxiv.org/abs/2311.05025*, 2024.
- Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 300–323, 2018.
- Ron Cole and Mark Fanty. ISOLET [Dataset]. UCI Machine Learning Repository, 1991. doi.org/10.24432/C51G69.
- Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In *Conference on Learning Theory*, pages 678–689, 2017.
- Arnak Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.
- Alain Durmus and Éric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.
- Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *Annals of Probability*, 47(4):1982–2010, 2019.
- Herbert Federer. *Geometric Measure Theory*. Springer, 1969.
- Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python Optimal Transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.
- James Foster. Approximating the signature of Brownian motion for high order SDE simulation. *arxiv.org/abs/2409.10118*, 2024.

- James Foster and Andraž Jelinčič. On the convergence of adaptive approximations for stochastic differential equations. *arxiv.org/abs/2311.14201*, 2024.
- James Foster, Terry Lyons, and Harald Oberhauser. An Optimal Polynomial Approximation of Brownian Motion. *SIAM Journal on Numerical Analysis*, 58(3):1393–1421, 2020.
- James Foster, Terry Lyons, and Harald Oberhauser. The shifted ODE method for underdamped Langevin MCMC. *arxiv.org/abs/2101.03446*, 2021.
- James Foster, Gonçalo dos Reis, and Calum Strange. High Order Splitting Methods for SDEs Satisfying a Commutativity Condition. *SIAM Journal on Numerical Analysis*, 62(1):500–532, 2024.
- Mike Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259–328, 2015.
- Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, 2010.
- Paul Hendricks. *titanic: Titanic Passenger Survival Data Set*, 2015. R package version 0.1.0. github.com/paulhendricks/titanic.
- Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.
- Zhengmian Hu, Feihu Huang, and Heng Huang. Optimal underdamped Langevin MCMC method. In *Advances in Neural Information Processing Systems*, pages 19363–19374, 2021.
- Silvana Ilie, Kenneth R. Jackson, and Wayne H. Enright. Adaptive time-stepping for the strong numerical solution of stochastic differential equations. *Numerical Algorithms*, 68(4):791–812, 2015.
- Andraž Jelinčič, James Foster, and Patrick Kidger. Single-seed generation of Brownian paths and integrals for adaptive and high order SDE solvers. *arxiv.org/abs/2405.06464*, 2024.
- Patrick Kidger. *On Neural Differential Equations*. DPhil thesis, University of Oxford, 2021. The DiffraX software package can be found at github.com/patrick-kidger/diffrax.
- Benedict Leimkuhler, Charles Matthews, and Gabriel Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA Journal of Numerical Analysis*, 36(1):13–79, 2016.
- Benedict Leimkuhler, Charles Matthews, and Jonathan Weare. Ensemble preconditioning for Markov chain Monte Carlo simulation. *Statistics and Computing*, 28:277–290, 2018.
- Benedict Leimkuhler, René Lohmann, and Peter Whalley. A Langevin sampling algorithm inspired by the Adam optimizer. *arxiv.org/abs/2504.18911*, 2025.

- Alix Leroy, Benedict Leimkuhler, Jonas Latz, and Desmond J Higham. Adaptive stepsize algorithms for Langevin dynamics. *SIAM Journal on Scientific Computing*, 46(6):3574–3598, 2024.
- Lingxiao Li, Qiang Liu, Anna Korba, Mikhail Yurochkin, and Justin Solomon. Sampling with Mollified Interaction Energy Descent. In *The Eleventh International Conference on Learning Representations*, 2023.
- Yi-An Ma, Tianqi Chen, and Emily B. Fox. A complete recipe for stochastic gradient MCMC. In *Advances in Neural Information Processing Systems*, pages 2917–2925, 2015.
- Stephan Mandt, Matthew D Hoffman, and David M Blei. Stochastic gradient descent as approximate Bayesian inference. *Journal of Machine Learning Research*, 18(1):4873–4907, 2017.
- Grigori N. Milstein and Michael V. Tretyakov. Quasi-symplectic methods for Langevin-type equations. *IMA Journal of Numerical Analysis*, 23(4):593–626, 2003.
- Pierre Monmarché. High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion. *Electronic Journal of Statistics*, 15(2):4117–4166, 2021.
- Richard W. Pastor. *The Molecular Dynamics of Liquid Crystals*. Springer, 1994.
- Richard W. Pastor, Bernard R. Brooks, and Attila Szabo. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics*, 65(6):1409–1419, 1988.
- Daniel Paulin and Peter Whalley. Correction to “Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations”. *Journal of Machine Learning Research*, 25(376):1–9, 2024.
- Grigorios A Pavliotis. *Stochastic Processes and Applications*. Springer, 2014.
- Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. arxiv.org/abs/1912.11554, 2019.
- Lionel Riou-Durand and Jure Vogrinc. Metropolis Adjusted Langevin Trajectories: a robust alternative to Hamiltonian Monte Carlo. arxiv.org/abs/2202.13230, 2022.
- Lionel Riou-Durand, Pavel Sountsov, Jure Vogrinc, Charles Margossian, and Sam Power. Adaptive Tuning for Metropolis Adjusted Langevin Trajectories. In *The 26th International Conference on Artificial Intelligence and Statistics*, 2023.
- Jesus María Sanz-Serna and Konstantinos C Zygalakis. Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations. *Journal of Machine Learning Research*, 22(1):11006–11042, 2021.
- Ruoqi Shen and Yin Tat Lee. The Randomized Midpoint Method for Log-Concave Sampling. In *Advances in Neural Information Processing Systems*, pages 2100–2111, 2019.
- Andrey Shkerin and Sergey Sibiryakov. Simple third order operator-splitting schemes for stochastic mechanics and field theory. arxiv.org/abs/2504.03844, 2025.

Zexi Song and Zhiqiang Tan. Hamiltonian Assisted Metropolis Sampling. *Journal of the American Statistical Association*, 118(542):1–50, 2021.

Paul Felix Valsecchi Oliva and Ömer Deniz Akyildiz. Kinetic Interacting Particle Langevin Monte Carlo. *arxiv.org/abs/2407.05790*, 2024.

Appendix A. Preliminary theorems

In this section, we detail several useful theorems that are crucial to proving our main results.

Theorem 14 (Global \mathbb{L}_p bounds of $\nabla f(x_t)$) *Let $\{(x_t, v_t)\}_{t \geq 0}$ denote the underdamped Langevin diffusion (2) with initial condition $(x_0, v_0) \sim \pi$, the unique stationary distribution of the diffusion process. Under the assumption that the gradient of f is M_1 -Lipschitz continuous then*

$$\|\nabla f(x_t)\|_{\mathbb{L}_{2p}} \leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{M_1 d}, \quad (27)$$

for all $p \geq 1$. Here, we use $!!$ to denote the standard double factorial for odd numbers.

Proof Our strategy is to prove (27) by induction for $d = 1$, and then extend it for $d \geq 1$. Recall that ULD has the stationary distribution

$$\pi(x, v) \propto \exp\left(-f(x) - \frac{1}{2u}\|v\|_2^2\right).$$

Therefore, if $(x_0, v_0) \sim \pi$ then $(x_t, v_t) \sim \pi$ for all $t \geq 0$.

For $d = 1$, we first consider the case where $p = 1$. Define

$$\hat{\pi}(x) = -f'(x)\pi(x),$$

and notice that $\hat{\pi}(x) = \frac{d}{dx}\pi(x)$. Then

$$\begin{aligned} \mathbb{E}[(f'(x))^2] &= \int_{-\infty}^{\infty} (f'(x))^2 \pi(x) dx = \int_{-\infty}^{\infty} f'(x) \left[f'(0) + \int_0^x f''(y) dy \right] \pi(x) dx \\ &= -f'(0) \int_{-\infty}^{\infty} \hat{\pi}(x) dx - \int_{-\infty}^{\infty} \int_0^x f''(y) \hat{\pi}(x) dy dx, \end{aligned}$$

by the fundamental theorem of calculus. For π to be well defined as a probability measure we require $\pi(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. Hence $\int_{-\infty}^{\infty} \hat{\pi}(x) dx = \lim_{x \rightarrow \infty} (\pi(x) - \pi(-x)) = 0$ and

$$\begin{aligned} \mathbb{E}[(f'(x))^2] &= - \int_{-\infty}^{\infty} \int_0^x f''(y) \hat{\pi}(x) dy dx \\ &= - \int_0^{\infty} \int_0^x f''(y) \hat{\pi}(x) dy dx + \int_{-\infty}^0 \int_x^0 f''(y) \hat{\pi}(x) dy dx \\ &= - \int_0^{\infty} \int_y^{\infty} f''(y) \hat{\pi}(x) dx dy + \int_{-\infty}^0 \int_{-\infty}^y f''(y) \hat{\pi}(x) dx dy \end{aligned}$$

$$= \int_0^\infty f''(y)\pi(y) dy + \int_{-\infty}^0 f''(y)\pi(y) dy = \int_{-\infty}^\infty f''(x)\pi(x) dx,$$

by Fubini's theorem and the properties of π , which in turn implies

$$\mathbb{E}[(f'(x))^2] \leq M_1,$$

as the derivative f' being M_1 -Lipschitz implies that f'' is bounded by M_1 almost everywhere.

For the inductive step, we will assume that for $p = n$

$$\mathbb{E}[(f'(x))^{2n}] = \int_{-\infty}^\infty (f'(x))^{2n} \pi(x) dx \leq (2n-1)!! M_1^n,$$

and for $p = n+1$, we define

$$\tilde{\pi}(x) = (f'(x))^{2n} \pi(x).$$

Notice that

$$\frac{d}{dx} \tilde{\pi}(x) = 2n f''(x) (f'(x))^{2n-1} \pi(x) - (f'(x))^{2n+1} \pi(x),$$

so rearranging gives

$$f'(x) \tilde{\pi}(x) = 2n f''(x) (f'(x))^{2n-1} \pi(x) - \tilde{\pi}'(x).$$

Therefore, from the definition of $\tilde{\pi}$ and the above identity, we have

$$\begin{aligned} & \mathbb{E}[(f'(x))^{2(n+1)}] \\ &= \int_{-\infty}^\infty f'(x) [f'(x) \tilde{\pi}(x)] dx \\ &= \int_{-\infty}^\infty f'(x) [2n f''(x) (f'(x))^{2n-1} \pi(x) - \tilde{\pi}'(x)] dx \\ &= 2n \int_{-\infty}^\infty f''(x) (f'(x))^{2n} \pi(x) dx - \int_{-\infty}^\infty f'(0) \tilde{\pi}'(x) dx - \int_{-\infty}^\infty \int_0^x f''(y) \tilde{\pi}'(x) dy dx, \end{aligned}$$

by the fundamental theorem of calculus.

Since $\int_{-\infty}^\infty \tilde{\pi}(x) dx$ is bounded (by the induction hypothesis) and $\tilde{\pi}$ is non-negative, it follows that $\tilde{\pi}(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. Therefore $\int_{-\infty}^\infty \tilde{\pi}'(x) dx = \lim_{x \rightarrow \infty} (\tilde{\pi}(x) - \tilde{\pi}(-x)) = 0$ and

$$\begin{aligned} & \mathbb{E}[(f'(x))^{2(n+1)}] \\ &= 2n \int_{-\infty}^\infty f''(x) (f'(x))^{2n} \pi(x) dx - \int_0^\infty \int_0^x f''(y) \tilde{\pi}'(x) dy dx + \int_{-\infty}^0 \int_x^0 f''(y) \tilde{\pi}'(x) dy dx \\ &= 2n \int_{-\infty}^\infty f''(x) (f'(x))^{2n} \pi(x) dx - \int_0^\infty \int_y^\infty f''(y) \tilde{\pi}'(x) dx dy + \int_{-\infty}^0 \int_{-\infty}^y f''(y) \tilde{\pi}'(x) dx dy \\ &= 2n \int_{-\infty}^\infty f''(x) (f'(x))^{2n} \pi(x) dx + \int_{-\infty}^\infty f''(x) \tilde{\pi}(x) dx, \end{aligned}$$

by Fubini's theorem and the property that $\tilde{\pi}(x) \rightarrow 0$ as $x \rightarrow \pm\infty$.

Hence, by the definition of $\tilde{\pi}$ and the fact that $|f''(x)| \leq M_1$ almost everywhere, we have

$$\int_{-\infty}^{\infty} (f'(x))^{2(n+1)} \pi(x) dx = (2n+1) \int_{-\infty}^{\infty} f''(x) (f'(x))^{2n} \pi(x) dx \leq (2n+1)!! M_1^{n+1}.$$

Since this bound is true for $p = 1$, it is also true for all $p \geq 1$ by induction.

For a general dimension $d \geq 1$, we can apply this bound to each partial derivative of f , as each marginal distribution of π is of the form $\exp(-g(x_i))$ where g is M_1 -Lipschitz. Then

$$\mathbb{E} \left[\|\nabla f(x)\|_2^{2p} \right] = \mathbb{E} \left[\left(\sum_{i=1}^d \left| \frac{\partial f}{\partial x_i} \right|^2 \right)^p \right] \leq d^{p-1} \mathbb{E} \left[\sum_{i=1}^d \left| \frac{\partial f}{\partial x_i} \right|^{2p} \right],$$

by the triangle inequality followed by Jensen's inequality (Foster et al., 2021, Theorem A.1). Therefore

$$\mathbb{E} \left[\|\nabla f(x)\|_2^{2p} \right] \leq d^{p-1} \sum_{i=1}^d \mathbb{E} \left[\left| \frac{\partial f}{\partial x_i} \right|^{2p} \right] \leq (2p-1)!! (M_1 d)^p,$$

as required. ■

Remark 15 *The above theorem can also be shown directly from the sub-Gaussianity of the stationary distribution. Therefore, the result holds more generally for an ergodic SDE initialised at its (sub-Gaussian) stationary distribution. Indeed, an alternative proof would involve using the strong concavity of π to deduce a logarithmic Sobolev inequality for π (Bakry et al., 2013), from which sub-Gaussianity follows via Herbst's argument.*

Remark 16 *A direct implication of Theorem 14 are the following bounds:*

$$\begin{aligned} \|\nabla f(x_t)\|_{\mathbb{L}_2} &\leq \sqrt{M_1 d}, \\ \|\nabla f(x_t)\|_{\mathbb{L}_4} &\leq \sqrt[4]{3} \sqrt{M_1 d}, \\ \|\nabla f(x_t)\|_{\mathbb{L}_6} &\leq \sqrt[6]{15} \sqrt{M_1 d}. \end{aligned} \tag{28}$$

Theorem 17 *Let $Z = (Z^{(1)}, \dots, Z^{(d)}) \in \mathbb{R}^d$ denote a centered d -dimensional Gaussian multivariate random variable such that $\text{Cov}(Z) = \sigma^2 I_d$ for some $\sigma \in \mathbb{R}$. Then*

$$\|Z\|_{\mathbb{L}_{2p}} \leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{\sigma^2 d}, \tag{29}$$

for all $p \geq 1$.

Proof First consider the case when $d = 1$ using the marginal $Z^{(i)} \sim \mathcal{N}(0, \sigma^2)$. It is a standard result (Bulmer, 1979, Chapter 7) that the $(2p)$ -th moment of $Z^{(i)}$ is

$$\mathbb{E}[(Z^{(i)})^{2p}] = (2p-1)!! \sigma^{2p}.$$

Then by Jensen's inequality (Foster et al., 2021, Theorem A.1), we have

$$\mathbb{E} \left[\left(\sum_{i=1}^d (Z^{(i)})^2 \right)^p \right] \leq d^{p-1} \sum_{i=1}^d \mathbb{E} \left[(Z^{(i)})^{2p} \right] \leq d^{p-1} \sum_{i=1}^d (2p-1)!! \sigma^{2p} = (2p-1)!! \sigma^{2p} d^p.$$

The result now directly follows since the left hand side is precisely equal to $\|Z\|_{\mathbb{L}_{2p}}^{2p}$. \blacksquare

Corollary 18 (Global \mathbb{L}_p bounds of v_t) *Let $\{(x_t, v_t)\}_{t \geq 0}$ be the underdamped Langevin diffusion with initial condition $(x_0, v_0) \sim \pi$, the stationary distribution of the process. Then*

$$\|v_t\|_{\mathbb{L}_{2p}} \leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{ud}, \quad (30)$$

for all $p \geq 1$.

Proof The stationary distribution of ULD (2) can be written explicitly as

$$\pi(x, v) \propto \exp \left(-f(x) - \frac{1}{2u} \|v\|_2^2 \right).$$

Thus, v_t has a centred d -dimensional multivariate Gaussian distribution with $\text{Var}(v_t^{(i)}) = u$. The result now immediately follows by Theorem 17. \blacksquare

Remark 19 *A direct implication of the Theorem 18 are the following bounds:*

$$\begin{aligned} \|v_t\|_{\mathbb{L}_2} &\leq \sqrt{ud}, \\ \|v_t\|_{\mathbb{L}_4} &\leq \sqrt[4]{3\sqrt{ud}}, \\ \|v_t\|_{\mathbb{L}_6} &\leq \sqrt[6]{15\sqrt{ud}}. \end{aligned} \quad (31)$$

Theorem 20 *Let $\{t_n\}_{n \geq 0}$ be a sequence of times with $t_0 = 0$, $t_{n+1} > t_n$ and $h_n = t_{n+1} - t_n$. Similar to the random variables given by (6), we define*

$$M_n = \frac{1}{h_n^3} \int_{t_n}^{t_{n+1}} \left(\frac{1}{10} h_n^2 - \frac{1}{2} (t - t_n) h_n + \frac{1}{2} (t - t_n)^2 \right) \left(W_{t_n, t} - \frac{t - t_n}{h_n} W_n \right) dt. \quad (32)$$

Then $M_n \sim \mathcal{N} \left(0, \frac{1}{100800} h_n I_d \right)$ is also independent of W_n , H_n and K_n . Furthermore,

$$\int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} W_{t_n, r_3} dr_3 dr_2 dr_1 = \frac{1}{24} h_n^3 W_n + \frac{3}{20} h_n^3 H_n + \frac{1}{2} h_n^3 K_n + h_n^3 M_n. \quad (33)$$

Proof One can derive (32) as a coefficient in a polynomial expansion of Brownian motion (see Foster et al. (2020, Theorem 2.2)). It is then straightforward to prove the identity (33) using integration by parts and Theorem 4. \blacksquare

Corollary 21 *Let W_n, H_n, K_n and M_n be the centred Gaussian random variables defined in equations (6) and (32). Then, for all $p \geq 1$, we have the bounds*

$$\begin{aligned} \|W_n\|_{\mathbb{L}_{2p}} &\leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{dh_n}, & \|H_n\|_{\mathbb{L}_{2p}} &\leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{\frac{1}{12}dh_n}, \\ \|K_n\|_{\mathbb{L}_{2p}} &\leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{\frac{1}{720}dh_n}, & \|M_n\|_{\mathbb{L}_{2p}} &\leq [(2p-1)!!]^{\frac{1}{2p}} \sqrt{\frac{1}{100800}dh_n}. \end{aligned} \quad (34)$$

Proof The result follows immediately from Theorem 17. ■

To help simplify the notation (especially for long calculations), we make the below definition.

Definition 22 *For all time intervals $[t_n, t_{n+1}]$, we define the following functions,*

$$\phi_0(x) = e^{-x\gamma h_n}, \quad \phi_1(x) = \frac{1 - e^{-x\gamma h_n}}{\gamma}, \quad \phi_2(x) = \frac{e^{-x\gamma h_n} + x\gamma h_n - 1}{\gamma^2}. \quad (35)$$

Theorem 23 *Consider a combination of the ϕ_i functions given by equation (35),*

$$p(h) = \sum_{i=1}^n \prod_{j=1}^m \phi_{a_j}(b_{i,j}),$$

where $a \in \{0, 1, 2\}^m$ and $b \in \{\lambda_+, \lambda_-, \frac{1}{3}\}^{n \times m}$. Then

$$|p(h)| \leq h^{\sum_{j=1}^m a_j} 2^{-\sum_{j=1}^m \mathbf{1}_{\{2\}}(a_j)} \sum_{i=1}^n \prod_{j=1}^m (b_{i,j})^{a_j}. \quad (36)$$

Proof Without loss of generality, we write

$$\phi_0(x, h) = e^{-x\gamma h}, \quad \phi_1(x, h) = \frac{1 - e^{-x\gamma h}}{\gamma}, \quad \phi_2(x, h) = \frac{e^{-x\gamma h} + x\gamma h - 1}{\gamma^2},$$

and note that

$$\phi_1(x, h) = \int_0^{xh} \phi_0(1, t) dt, \quad \phi_2(x, h) = \int_0^{xh} \int_0^{r_1} \phi_0(1, r_2) dr_2 dr_1.$$

Since $|\phi_0(x, h)| \leq 1$ for all $x, h \geq 0$, then

$$\begin{aligned} |\phi_1(x, h)| &= \left| \int_0^{xh} \phi_0(1, t) dt \right| \leq \int_0^{xh} |\phi_0(1, t)| dt \leq \int_0^{xh} 1 dt = xh, \\ |\phi_2(x, h)| &= \left| \int_0^{xh} \int_0^{r_1} \phi_0(1, r_2) dr_2 dr_1 \right| \leq \int_0^{xh} \int_0^{r_1} |\phi_0(1, r_2)| dr_2 dr_1 \leq \frac{1}{2} x^2 h^2. \end{aligned}$$

Therefore, by the triangle inequality

$$\left| \sum_{i=1}^n \prod_{j=1}^m \phi_{a_j}(b_{i,j}) \right| \leq \sum_{i=1}^n \left| \prod_{j=1}^m \phi_{a_j}(b_{i,j}) \right| = \sum_{i=1}^n \prod_{j=1}^m |\phi_{a_j}(b_{i,j})|.$$

The result now follows after taking the inequalities of the ϕ_i functions and rearranging. ■

To end this section, we present a few useful results for estimating terms in Taylor expansions.

Theorem 24 *Under the assumptions (A.1) and (A.2), we have*

$$m\|v\|_{\mathbb{L}_p} \leq \|\nabla^2 f(x)v\|_{\mathbb{L}_p} \leq M_1\|v\|_{\mathbb{L}_p}, \quad (37)$$

for all $x, v \in \mathbb{R}^d$ and $p \geq 1$. Furthermore, for all $x, v_1, v_2 \in \mathbb{R}^d$ and $p \geq 1$,

$$\|\nabla^2 f(x)v_1 - \nabla^2 f(x)v_2\|_{\mathbb{L}_p} \leq M_1\|v_1 - v_2\|_{\mathbb{L}_p}. \quad (38)$$

Under the additional assumption (A.3)

$$\|\nabla^2 f(x_1)v - \nabla^2 f(x_2)v\|_{\mathbb{L}_p} \leq M_2\|x_1 - x_2\|_{\mathbb{L}_{2p}}\|v\|_{\mathbb{L}_{2p}}, \quad (39)$$

for all $x_1, x_2, v \in \mathbb{R}^d$ and $p \geq 1$. In addition, we have

$$\|\nabla^3 f(x)(v_1, v_2)\|_{\mathbb{L}_p} \leq M_2\|v_1\|_{\mathbb{L}_{2p}}\|v_2\|_{\mathbb{L}_{2p}}, \quad (40)$$

and

$$\|\nabla^3 f(x)(v_1, v_2) - \nabla^3 f(x)(v_1, v_3)\|_{\mathbb{L}_p} \leq M_2\|v_1\|_{\mathbb{L}_{2p}}\|v_2 - v_3\|_{\mathbb{L}_{2p}}, \quad (41)$$

for all $x, v_1, v_2, v_3 \in \mathbb{R}^d$ and $p \geq 1$. Finally, under the additional assumption (A.4)

$$\|\nabla^3 f(x_1)(v_1, v_2) - \nabla^3 f(x_2)(v_1, v_2)\|_{\mathbb{L}_p} \leq M_3\|x_1 - x_2\|_{\mathbb{L}_{3p}}\|v_1\|_{\mathbb{L}_{3p}}\|v_2\|_{\mathbb{L}_{3p}}, \quad (42)$$

for all $x_1, x_2, v_1, v_2 \in \mathbb{R}^d$ and $p \geq 1$.

Proof These inequalities follow from the definitions of \mathbb{L}_p norm and Lipschitz continuity, along with an application of Hölder's inequality. ■

Theorem 25 *Let $c > 0$ be a fixed constant. Then for vectors $x, y \in \mathbb{R}^d$, we have*

$$\langle x, y \rangle \leq \frac{c}{2}\|x\|_2^2 + \frac{1}{2c}\|y\|_2^2. \quad (43)$$

Proof Using the Cauchy-Schwarz and Young inequalities, we obtain

$$\langle x, y \rangle = \left\langle \sqrt{c}x, \frac{1}{\sqrt{c}}y \right\rangle \leq \|\sqrt{c}x\|_2 \left\| \frac{1}{\sqrt{c}}y \right\|_2 \leq \frac{1}{2} \left(\|\sqrt{c}x\|_2^2 + \left\| \frac{1}{\sqrt{c}}y \right\|_2^2 \right) = \frac{1}{2} \left(c\|x\|_2^2 + \frac{1}{c}\|y\|_2^2 \right),$$

as required. ■

Appendix B. Exponential contractivity of the QUICSORT method

In this section, we will prove Theorem 8, which gives the contractivity of the QUICSORT method. Our first step will be to write the QUICSORT method (3) as the sum of two terms,

$$X_{n+1} = \tilde{X}_{n+1} + \bar{X}_n, \quad V_{n+1} = \tilde{V}_{n+1} + \bar{V}_n, \quad (44)$$

where

$$\begin{aligned} \tilde{V}_n^{(1)} &:= V_n + \sigma(H_n + 6K_n), \\ \tilde{X}_n^{(1)} &:= X_n + \frac{1 - e^{-\lambda-\gamma h_n}}{\gamma} \tilde{V}_n^{(1)} + \frac{e^{-\lambda-\gamma h_n} + \lambda_- \gamma h_n - 1}{\gamma^2} C_n, \\ \tilde{V}_n^{(2)} &:= e^{-\lambda+\gamma h_n} \tilde{V}_n^{(1)} - \frac{1}{2} e^{-\frac{1}{3}\gamma h_n} u \nabla f(\tilde{X}_n^{(1)}) h_n + \frac{1 - e^{-\lambda+\gamma h_n}}{\gamma} C_n, \\ \tilde{X}_n^{(2)} &:= X_n + \frac{1 - e^{-\lambda+\gamma h_n}}{\gamma} \tilde{V}_n^{(1)} - \frac{1 - e^{-\frac{1}{3}\gamma h_n}}{\gamma} u \nabla f(\tilde{X}_n^{(1)}) h_n + \frac{e^{-\lambda+\gamma h_n} + \lambda_+ \gamma h_n - 1}{\gamma^2} C_n, \\ \tilde{V}_n^{(3)} &:= e^{-\lambda-\gamma h_n} \tilde{V}_n^{(2)} - \frac{1}{2} e^{-\lambda-\gamma h_n} u \nabla f(\tilde{X}_n^{(2)}) h_n + \frac{1 - e^{-\lambda-\gamma h_n}}{\gamma} C_n, \\ \tilde{X}_{n+1} &:= \tilde{X}_n^{(2)} - \frac{1 - e^{-\lambda-\gamma h_n}}{2\gamma} u \nabla f(\tilde{X}_n^{(2)}) h_n + \frac{1 - e^{-\lambda-\gamma h_n}}{\gamma} \tilde{V}_n^{(2)} + \frac{e^{-\lambda-\gamma h_n} + \lambda_- \gamma h - 1}{\gamma^2} C_n, \\ \tilde{V}_{n+1} &:= \tilde{V}_n^{(3)} + \sigma(H_n - 6K_n), \end{aligned}$$

and

$$\begin{aligned} \bar{X}_n &:= \frac{1}{2\gamma} \left(1 - e^{-\frac{1}{3}\gamma h_n} - e^{-(\frac{1}{3}+\lambda_-)\gamma h_n} + e^{-\lambda+\gamma h_n} \right) u \nabla f(\tilde{X}_n^{(1)}) h_n, \\ \bar{V}_n &:= \frac{1}{2} \left(e^{-(\frac{1}{3}+\lambda_-)\gamma h_n} - e^{-\lambda+\gamma h_n} \right) u \nabla f(\tilde{X}_n^{(1)}) h_n. \end{aligned}$$

The procedure for showing that the QUICSORT approximation is contractive is as follows. Since proving contractivity for a map with multiple gradient evaluations can be difficult, we will split the entire step $(X_n, V_n) \rightarrow (X_{n+1}, V_{n+1})$ into the following three smaller mappings:

$$(X_n, V_n) \rightarrow (\tilde{X}_n^{(2)}, \tilde{V}_n^{(2)}), \quad (\tilde{X}_n^{(2)}, \tilde{V}_n^{(2)}) \rightarrow (\tilde{X}_{n+1}, \tilde{V}_{n+1}), \quad (\tilde{X}_{n+1}, \tilde{V}_{n+1}) \rightarrow (X_{n+1}, V_{n+1}).$$

The first two maps are designed to resemble the UBU and Exponential Euler schemes (Sanz-Serna and Zygalkakis, 2021) and so, by similar techniques, can be shown to be contractive. The final map is simply adding the correction term (\bar{X}_n, \bar{V}_n) , which we will show is small.

Theorem 26 *Consider the two approximations (X, V) and (Y, U) under the mappings $(X_n, V_n) \rightarrow (\tilde{X}_n^{(2)}, \tilde{V}_n^{(2)})$ and $(Y_n, U_n) \rightarrow (\tilde{Y}_n^{(2)}, \tilde{U}_n^{(2)})$ driven by the same Brownian motion but evolved from initial conditions (X_n, V_n) and (Y_n, U_n) . We define the transformations,*

$$\begin{aligned} \begin{pmatrix} W_n \\ Z_n \end{pmatrix} &:= \begin{pmatrix} (wX_n + V_n) - (wY_n + U_n) \\ (zX_n + V_n) - (zY_n + U_n) \end{pmatrix}, \\ \begin{pmatrix} W_n^{(1)} \\ Z_n^{(1)} \end{pmatrix} &:= \begin{pmatrix} (wX_n^{(2)} + V_n^{(2)}) - (wY_n^{(2)} + U_n^{(2)}) \\ (zX_n^{(2)} + V_n^{(2)}) - (zY_n^{(2)} + U_n^{(2)}) \end{pmatrix}, \end{aligned} \quad (45)$$

where $w \in [0, \frac{1}{2}\gamma)$ and $z = \gamma - w$. Then, under assumptions (A.1) and (A.2), we have

$$(\|W_n^{(1)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(1)}\|_{\mathbb{L}_2}^2) \leq (1 - \alpha_1 h_n)(\|W_n\|_{\mathbb{L}_2}^2 + \|Z_n\|_{\mathbb{L}_2}^2), \quad (46)$$

for all $0 < h_n \leq 1$ and $n \geq 0$ where

$$\alpha_1 = \frac{(2\lambda_+ z^2 - uM_1) \wedge (um - 2\lambda_+ w^2)}{z - w} - K_1 h_n, \quad (47)$$

for some constant $K_1 > 0$ not depending on h_n .

Proof We can unfold the mappings and then simplify notation using the ϕ_i functions (35). The noise terms cancel out due to the synchronous coupling, which first yields

$$\begin{aligned} \tilde{V}_n^{(1)} - \tilde{U}_n^{(1)} &= V_n - U_n, \\ \tilde{X}_n^{(1)} - \tilde{Y}_n^{(1)} &= X_n - Y_n + \frac{1 - e^{-\lambda_- \gamma h_n}}{\gamma} (V_n - U_n) \\ &= X_n - Y_n + \phi_1(\lambda_-)(V_n - U_n). \end{aligned}$$

Since we assume that f is twice differentiable, we can use Taylor's theorem to define⁴

$$\begin{aligned} \nabla f(\tilde{X}_n^{(1)}) - \nabla f(\tilde{Y}_n^{(1)}) &= \underbrace{\int_0^1 \nabla^2 f(\tilde{X}_n^{(1)} + r(\tilde{Y}_n^{(1)} - \tilde{X}_n^{(1)})) dr}_{=: \mathcal{H}_n^{(1)}} (\tilde{X}_n^{(1)} - \tilde{Y}_n^{(1)}) \\ &= \mathcal{H}_n^{(1)} [(X_n - Y_n) + \phi_1(\lambda_-)(V_n - U_n)], \end{aligned}$$

where $\mathcal{H}_n^{(1)} \in \mathbb{R}^{d \times d}$ is a symmetric positive semidefinite matrix which, from the assumptions (A.1) and (A.2), additionally satisfies $mI_d \preceq \mathcal{H}_n^{(1)} \preceq M_1 I_d$.

Then the next stage in the mappings yields following differences,

$$\begin{aligned} \tilde{V}_n^{(2)} - \tilde{U}_n^{(2)} &= e^{-\lambda_+ \gamma h_n} (V_n - U_n) - \frac{1}{2} e^{-\frac{1}{3} \gamma h_n} u h_n (\nabla f(\tilde{X}_n^{(1)}) - \nabla f(\tilde{Y}_n^{(1)})) \\ &= \left[-\frac{1}{2} u h_n \phi_0 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)} \right] (X_n - Y_n) + \left[\phi_0(\lambda_+) I_d - \frac{1}{2} u h_n \phi_0 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)} \right] (V_n - U_n), \end{aligned}$$

$$\begin{aligned} \tilde{X}_n^{(2)} - \tilde{Y}_n^{(2)} &= (X_n - Y_n) + \frac{1 - e^{-\lambda_+ \gamma h_n}}{\gamma} (V_n - U_n) - \frac{1 - e^{-\frac{1}{3} \gamma h_n}}{\gamma} u h_n (\nabla f(\tilde{X}_n^{(1)}) - \nabla f(\tilde{Y}_n^{(1)})) \\ &= \left[I_d - u h_n \phi_1 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)} \right] (X_n - Y_n) + \left[\phi_1(\lambda_+) I_d - u h_n \phi_1 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)} \right] (V_n - U_n). \end{aligned}$$

4. The following definition is still valid even if we do not assume that the potential f is twice differentiable. By Rademacher's theorem (Federer, 1969, Theorem 3.1.6) since ∇f is Lipschitz continuous, then ∇f is almost everywhere differentiable. Therefore, as the probability of hitting a null set is 0, we can infer that the potential function's second derivative is sufficiently well defined to be applied in our setting.

From the definitions of W_n and Z_n , we can write

$$X_n - Y_n = \frac{Z_n - W_n}{z - w}, \quad V_n - U_n = \frac{zW_n - wZ_n}{z - w}.$$

Thus, $W_n^{(1)}$ can be expressed as

$$\begin{aligned} W_n^{(1)} &= w(\tilde{X}_n^{(2)} - \tilde{Y}_n^{(2)}) + (\tilde{V}_n^{(2)} - \tilde{U}_n^{(2)}) \\ &= \left(w \left[I_d - uh_n \phi_1 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)} \right] + \left[-\frac{1}{2} uh_n \phi_0 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)} \right] \right) (X_n - Y_n) \\ &\quad + w \left[\phi_1(\lambda_+) I_d - uh_n \phi_1 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)} \right] (V_n - U_n) \\ &\quad + \left[\phi_0(\lambda_+) I_d - \frac{1}{2} uh_n \phi_0 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)} \right] (V_n - U_n) \\ &= \frac{1}{z - w} \left(\left(w \left[I_d - uh_n \phi_1 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)} \right] + \left[-\frac{1}{2} uh_n \phi_0 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)} \right] \right) (Z_n - W_n) \right. \\ &\quad \left. + w \left[\phi_1(\lambda_+) I_d - uh_n \phi_1 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)} \right] (zW_n - wZ_n) \right. \\ &\quad \left. + \left[\phi_0(\lambda_+) I_d - \frac{1}{2} uh_n \phi_0 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)} \right] (zW_n - wZ_n) \right) \\ &= \frac{1}{z - w} \left[(-wA - B + wzC + zD)W_n + (wA + B - w^2C - wD)Z_n \right], \end{aligned}$$

and similarly

$$Z_n^{(1)} = \frac{1}{z - w} \left[(-zA - B + z^2C + zD)W_n + (zA + B - wzC - wD)Z_n \right],$$

where

$$\begin{aligned} A &= I_d - uh_n \phi_1 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)}, & B &= -\frac{1}{2} uh_n \phi_0 \left(\frac{1}{3} \right) \mathcal{H}_n^{(1)}, \\ C &= \phi_1(\lambda_+) I_d - uh_n \phi_1 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)}, & D &= \phi_0(\lambda_+) I_d - \frac{1}{2} uh_n \phi_0 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \mathcal{H}_n^{(1)}. \end{aligned}$$

Note the matrices A, B, C, D are not only symmetric but also commute with one another. By expanding $\|W_n^{(1)}\|_2^2$ and $\|Z_n^{(1)}\|_2^2$, we obtain

$$\|W_n^{(1)}\|_2^2 + \|Z_n^{(1)}\|_2^2 = \frac{1}{(z - w)^2} (W_n^\top E W_n + Z_n^\top F Z_n + 2W_n^\top G Z_n), \quad (48)$$

where

$$\begin{aligned} E &= (w^2 + z^2)A^2 + 2B^2 + z^2(w^2 + z^2)C^2 + 2z^2D^2 + 2\gamma AB - 2z(w^2 + z^2)AC \\ &\quad - 2\gamma zAD - 2\gamma zBC - 4zBD + 2\gamma z^2CD, \\ F &= (w^2 + z^2)A^2 + 2B^2 + w^2(w^2 + z^2)C^2 + 2w^2D^2 + 2\gamma AB - 2w(w^2 + z^2)AC \\ &\quad - 2w\gamma AD - 2\gamma wBC - 4wBD + 2\gamma w^2CD, \\ G &= -(w^2 + z^2)A^2 - 2B^2 - zw(w^2 + z^2)C^2 - 2zwD^2 - 2\gamma AB + \gamma(w^2 + z^2)AC \\ &\quad + \gamma^2 AD + \gamma^2 BC + 2\gamma BD - 2\gamma wzCD. \end{aligned}$$

Whilst we could bound the terms involving A , B , C and D by applying the inequality (36), since $-\phi_i(x) \leq |\phi_i(x)|$ for all $x \in \{\lambda_+, \lambda_-, \frac{1}{3}\}$ and $i \in \{0, 1\}$ this would produce too high an upper bound in order to prove the desired result. Instead, we will use the bounds below,

$$-\phi_0(x) \leq -1 + x\gamma h_n, \quad -\phi_1(x) \leq -xh_n + \frac{1}{2}\gamma x^2 h_n^2.$$

In the following, we will also use

$$-h_n^c \sum_{i=1}^m \phi_{a_i}(b_i) \leq 0,$$

for terms with $c + \sum_{i=1}^m a_i \geq 2$ as they do not significantly impact the contraction rate α_1 and all the ϕ functions are bounded below by 0. Using the Loewner order \preceq and the commutativity of the matrices A, B, C and D , we can bound the matrices E, F and G as

$$\begin{aligned} E &\preceq (w^2 + z^2) \left(I_d + \frac{1}{9} u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2 \left(\frac{1}{4} u^2 h_n^2 (\mathcal{H}_n^{(1)})^2 \right) + z^2 (w^2 + z^2) \left(\lambda_+^2 h_n^2 I_d + \frac{2}{9} \lambda_-^2 u^2 h_n^6 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2z^2 \left(I_d - 2\lambda_+ \gamma h_n I_d + 2\lambda_+^2 \gamma^2 h_n^2 I_d + \frac{1}{4} \lambda_-^2 u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2\gamma \left(-\frac{1}{2} u h_n \mathcal{H}_n^{(1)} + \frac{1}{6} \gamma u h_n^2 \mathcal{H}_n^{(1)} + \frac{1}{6} u^2 h_n^3 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2z(w^2 + z^2) \left(\lambda_+ h_n I_d + \frac{1}{2} \gamma \lambda_+^2 h_n^2 I_d + \frac{1}{3} u h_n^3 \mathcal{H}_n^{(1)} \right) \\ &\quad + 2\gamma z \left(I_d - \lambda_+ \gamma h_n I_d + \frac{1}{2} \lambda_+^2 \gamma^2 h_n^2 I_d + \frac{1}{6} \lambda_- u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2\gamma z \left(\frac{1}{2} \lambda_+ u h_n^2 \mathcal{H}_n^{(1)} \right) + 4z \left(\frac{1}{2} u h_n \mathcal{H}_n^{(1)} \right) + 2\gamma z^2 \left(\lambda_+ h_n I_d + \frac{1}{6} \lambda_-^2 u^2 h_n^5 (\mathcal{H}_n^{(1)})^2 \right) \\ &\preceq \left((z-w)^2 + (z-w)h_n(uM_1 - 2\lambda_+ z^2) + h_n^2(\gamma z + 3\gamma^2 z^2 + z^2 w^2 + z^4) \right. \\ &\quad \left. + h_n^2[(z + \gamma w + 2\gamma z)uM_1 + (1 + \gamma + z^2 + w^2 + z^2 w^2 + z^4)u^2 M_1^2] \right) I_d \\ &= \left((z-w)^2 + (z-w)h_n(uM_1 - 2\lambda_+ z^2) + E_+ h_n^2 \right) I_d, \\ F &\preceq (w^2 + z^2) \left(I_d + \frac{1}{9} u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) + 2 \left(\frac{1}{4} u^2 h_n^2 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + w^2 (w^2 + z^2) \left(\lambda_+^2 h_n^2 I_d + \frac{2}{9} \lambda_-^2 u^2 h_n^6 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2w^2 \left(I_d - 2\lambda_+ \gamma h_n I_d + 2\lambda_+^2 \gamma^2 h_n^2 I_d + \frac{1}{4} \lambda_-^2 u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) \\ &\quad + 2\gamma \left(-\frac{1}{2} u h_n \mathcal{H}_n^{(1)} + \frac{1}{6} \gamma u h_n^2 \mathcal{H}_n^{(1)} + \frac{1}{6} u^2 h_n^3 (\mathcal{H}_n^{(1)})^2 \right) \end{aligned}$$

$$\begin{aligned}
 & + 2w(w^2 + z^2) \left(\lambda_+ h_n I_d + \frac{1}{2} \gamma \lambda_+^2 h_n^2 I_d + \frac{1}{3} u h_n^3 \mathcal{H}_n^{(1)} \right) \\
 & + 2\gamma w \left(I_d - \lambda_+ \gamma h_n I_d + \frac{1}{2} \lambda_+^2 \gamma^2 h_n^2 I_d + \frac{1}{6} \lambda_- u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) \\
 & + 2\gamma w \left(\frac{1}{2} \lambda_+ u h_n^2 \mathcal{H}_n^{(1)} \right) + 4w \left(\frac{1}{2} u h_n \mathcal{H}_n^{(1)} \right) + 2\gamma w^2 \left(\lambda_+ h_n I_d + \frac{1}{6} \lambda_-^2 u^2 h_n^5 (\mathcal{H}_n^{(1)})^2 \right) \\
 & \preceq \left((z-w)^2 + (z-w) h_n (2\lambda_+ w^2 - um) + h_n^2 (\gamma w + 3\gamma^2 w^2 + z^2 w^2 + w^4) \right. \\
 & \quad \left. + h_n^2 [(w + \gamma z + 2\gamma w) u M_1 + (1 + \gamma + z^2 + w^2 + z^2 w^2 + w^4) u^2 M_1^2] \right) I_d \\
 & = \left((z-w)^2 + (z-w) h_n (2\lambda_+ w^2 - um) + F_+ h_n^2 \right) I_d, \\
 G & \preceq (w^2 + z^2) \left(-I_d + \frac{2}{3} u h_n^2 \mathcal{H}_n^{(1)} \right) \\
 & + zw(w^2 + z^2) \left(\lambda_+^2 h_n^2 I_d - \gamma \lambda_+^3 h_n^3 I_d + \frac{1}{4} \gamma^2 \lambda_-^4 h_n^4 I_d + \frac{2}{3} \lambda_+ \lambda_- u h_n^4 \mathcal{H}_n^{(1)} \right) \\
 & + 2zw \left(-I_d + 2\lambda_+ \gamma h_n I_d + \lambda_- u h_n^2 \mathcal{H}_n^{(1)} \right) \\
 & + \gamma u h_n \mathcal{H}_n^{(1)} + \gamma (w^2 + z^2) \left(\lambda_+ h_n I_d + \frac{1}{9} \lambda_- u^2 h_n^5 (\mathcal{H}_n^{(1)})^2 \right) \\
 & + \gamma^2 \left(I_d - \lambda_+ \gamma h_n I_d + \frac{1}{2} \lambda_+^2 \gamma^2 h_n^2 I_d + \frac{1}{6} \lambda_- u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) \\
 & + \gamma^2 \left(\frac{1}{6} \lambda_- u^2 h_n^4 (\mathcal{H}_n^{(1)})^2 \right) + \gamma \left(-u h_n \mathcal{H}_n^{(1)} + \left(\frac{1}{3} + \lambda_+ \right) \gamma u h_n^2 \mathcal{H}_n^{(1)} \right) \\
 & + 2\gamma wz \left(-\lambda_+ h_n I_d + 2\gamma \lambda_+^2 h_n^2 I_d + \frac{1}{2} \lambda_+ \lambda_- u h_n^3 \mathcal{H}_n^{(1)} + \frac{1}{3} \lambda_- u h_n^3 \mathcal{H}_n^{(1)} \right) \\
 & + \frac{1}{2} \lambda_- \gamma u^2 h_n^3 (\mathcal{H}_n^{(1)})^2 \\
 & \preceq h_n^2 \left[z^3 w + z w^3 + 3\gamma^2 z w + \gamma^2 z^3 w + \gamma^2 z w^3 + \gamma^4 + (z w + z^2 + w^2 + 2\gamma^2) u M_1 \right. \\
 & \quad \left. + (\gamma z w + z^3 w + z w^3) u M_1 + (1 + \gamma + \gamma^2 + \gamma z^2 + \gamma w^2) u^2 M_1^2 \right] I_d \\
 & = G_+ h_n^2 I_d,
 \end{aligned}$$

where we have used the assumption that $h_n \leq 1$ and also bounded λ_+ and λ_- for readability. Next, we will establish an $O(h_n^2)$ upper bound for $W_n^\top G Z_n$ in terms of $\|W_n\|_{\mathbb{L}_2}^2$ and $\|Z_n\|_{\mathbb{L}_2}^2$.

$$\begin{aligned}
 4W_n^\top G Z_n & = (W_n + Z_n)^\top G (W_n + Z_n) - (W_n - Z_n)^\top G (W_n - Z_n) \\
 & \leq |(W_n + Z_n)^\top G (W_n + Z_n)| + |(W_n - Z_n)^\top G (W_n - Z_n)| \\
 & \leq G_+ h_n^2 \|W_n + Z_n\|_2^2 + G_+ h_n^2 \|W_n - Z_n\|_2^2,
 \end{aligned}$$

which, by the triangle and Young's inequalities, implies that

$$W_n^\top G Z_n \leq G_+ h_n^2 \left(\|W_n\|_2^2 + \|Z_n\|_2^2 \right).$$

Therefore, putting this all together gives

$$\begin{aligned}
 & \|W_n^{(1)}\|_2^2 + \|Z_n^{(1)}\|_2^2 \\
 &= \frac{1}{(z-w)^2} \left(W_n^\top E W_n + Z_n^\top F Z_n + 2W_n^\top G Z_n \right) \\
 &\leq \|W_n\|_2^2 + \|Z_n\|_2^2 + \frac{h_n}{z-w} \left((uM_1 - 2\lambda_+ z^2) \|W_n\|_2^2 + (2\lambda_+ w^2 - um) \|Z_n\|_2^2 \right) \\
 &\quad + \frac{h_n^2}{(z-w)^2} \left(E_+ \|W_n\|_2^2 + F_+ \|Z_n\|_2^2 + 2G_+ (\|W_n\|_2^2 + \|Z_n\|_2^2) \right) \\
 &\leq \left(1 + \frac{h_n}{z-w} \max((uM_1 - 2\lambda_+ z^2), (2\lambda_+ w^2 - um)) \right) \\
 &\quad + \frac{h_n^2}{(z-w)^2} (\max(E_+, F_+) + 2G_+) (\|W_n\|_2^2 + \|Z_n\|_2^2) \\
 &= (1 - \alpha_1 h_n) (\|W_n\|_2^2 + \|Z_n\|_2^2).
 \end{aligned}$$

The result now follows by taking expectations of both sides. \blacksquare

Similarly, we can show that the second map $(\tilde{X}_n^{(2)}, \tilde{V}_n^{(2)}) \rightarrow (\tilde{X}_{n+1}, \tilde{V}_{n+1})$ is also contractive.

Theorem 27 *Consider the two approximations (X, V) and (Y, U) under the mappings $(\tilde{X}_n^{(2)}, \tilde{V}_n^{(2)}) \rightarrow (\tilde{X}_{n+1}, \tilde{V}_{n+1})$ and $(\tilde{Y}_n^{(2)}, \tilde{U}_n^{(2)}) \rightarrow (\tilde{Y}_{n+1}, \tilde{U}_{n+1})$ that are driven by the same Brownian motion but are evolved from the initial conditions $(\tilde{X}_n^{(2)}, \tilde{V}_n^{(2)})$ and $(\tilde{Y}_n^{(2)}, \tilde{U}_n^{(2)})$. Just as before, we define the following coordinate transformations,*

$$\begin{aligned}
 \begin{pmatrix} W_n^{(1)} \\ Z_n^{(1)} \end{pmatrix} &= \begin{pmatrix} (wX_n^{(2)} + V_n^{(2)}) - (wY_n^{(2)} + U_n^{(2)}) \\ (zX_n^{(2)} + V_n^{(2)}) - (zY_n^{(2)} + U_n^{(2)}) \end{pmatrix}, \\
 \begin{pmatrix} W_n^{(2)} \\ Z_n^{(2)} \end{pmatrix} &= \begin{pmatrix} (w\tilde{X}_{n+1} + \tilde{V}_{n+1}) - (w\tilde{Y}_{n+1} + \tilde{U}_{n+1}) \\ (z\tilde{X}_{n+1} + \tilde{V}_{n+1}) - (z\tilde{Y}_{n+1} + \tilde{U}_{n+1}) \end{pmatrix},
 \end{aligned} \tag{49}$$

where $w \in [0, \frac{1}{2}\gamma)$ and $z = \gamma - w$. Then, under assumptions (A.1) and (A.2), we have

$$(\|W_n^{(2)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(2)}\|_{\mathbb{L}_2}^2) \leq (1 - \alpha_2 h_n) (\|W_n^{(1)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(1)}\|_{\mathbb{L}_2}^2), \tag{50}$$

for all $0 < h_n \leq 1$ and $n \geq 0$ where

$$\alpha_2 = \frac{(2\lambda_- z^2 - uM_1) \wedge (um - 2\lambda_- w^2)}{z-w} - K_2 h_n, \tag{51}$$

for some constant $K_2 > 0$.

Proof The proof strategy is identical to that of Theorem 26. The main difference is simply the values of the constants E_+ , F_+ , and G_+ . Specifically, these constants are given by

$$E_+ = \gamma^2 z^2 + \gamma z^3 + \gamma z w^2 + z^2 w^2 + z^4 + (\gamma + \gamma z + z w^2 + z^3) u M_1 + (1 + \gamma) u^2 M_1^2,$$

$$\begin{aligned}
 F_+ &= \gamma^2 w^2 + \gamma w^3 + \gamma z^2 w + z^2 w^2 + w^4 + (\gamma + \gamma w + z^2 w + w^3) u M_1 + (1 + \gamma) u^2 M_1^2, \\
 G_+ &= \gamma^2 z w + \gamma^4 + (\gamma^2 + z^2 + w^2) u M_1.
 \end{aligned}$$

■

Next, we will show that the third mapping $(\tilde{X}_{n+1}, \tilde{V}_{n+1}) \rightarrow (X_{n+1}, V_{n+1})$ is always small.

Theorem 28 *Consider the two approximations (X, V) and (Y, U) under the mappings $(\tilde{X}_{n+1}, \tilde{V}_{n+1}) \rightarrow (X_{n+1}, V_{n+1})$ and $(\tilde{Y}_{n+1}, \tilde{U}_{n+1}) \rightarrow (Y_{n+1}, U_{n+1})$ that are driven by the same Brownian motion but evolved from initial conditions $(\tilde{X}_{n+1}, \tilde{V}_{n+1})$ and $(\tilde{Y}_{n+1}, \tilde{U}_{n+1})$. Just as before, we define the following coordinate transformations,*

$$\begin{aligned}
 \begin{pmatrix} W_n \\ Z_n \end{pmatrix} &= \begin{pmatrix} (wX_n + V_n) - (wY_n + U_n) \\ (zX_n + V_n) - (zY_n + U_n) \end{pmatrix}, \\
 \begin{pmatrix} W_n^{(3)} \\ Z_n^{(3)} \end{pmatrix} &= \begin{pmatrix} (w\bar{X}_n + \bar{V}_n) - (w\bar{Y}_n + \bar{U}_n) \\ (z\bar{X}_n + \bar{V}_n) - (z\bar{Y}_n + \bar{U}_n) \end{pmatrix},
 \end{aligned} \tag{52}$$

where $w \in [0, \frac{1}{2}\gamma)$ and $z = \gamma - w$. Then, under assumptions (A.1) and (A.2), we have

$$(\|W_n^{(3)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(3)}\|_{\mathbb{L}_2}^2) \leq K_3 h_n^4 (\|W_n\|_{\mathbb{L}_2}^2 + \|Z_n\|_{\mathbb{L}_2}^2), \tag{53}$$

for all $0 < h_n \leq 1$ and $n \geq 0$ where $K_3 > 0$ is a constant not depending on h_n .

Proof Again, the proof strategy is nearly identical to that of Theorem 26. Here, we use

$$E \leq E_+ h_n^4, \quad \text{and} \quad F \leq F_+ h_n^4,$$

where

$$\begin{aligned}
 E_+ &\leq (1 + \gamma + z^2 + w^2 + \gamma z^2 + z^2 w^2 + z^4) u^2 M_1^2, \\
 F_+ &\leq (1 + \gamma + z^2 + w^2 + \gamma w^2 + z^2 w^2 + w^4) u^2 M_1^2,
 \end{aligned}$$

and

$$G_+ \leq (\gamma + \gamma^2 + \gamma z^2 + \gamma w^2) u^2 M_1^2 h_n^2.$$

■

Finally, we are now in a position to prove our main contractivity result (Theorem 8).

Theorem 29 (Contractivity of the QUICSORT method (1)) *Let (X, V) and (Y, U) denote two numerical approximations defined by the mappings $(X_n, V_n) \rightarrow (X_{n+1}, V_{n+1})$ and $(Y_n, U_n) \rightarrow (Y_{n+1}, U_{n+1})$ which are driven by the same Brownian motion but evolved from initial conditions (X_n, V_n) and (Y_n, U_n) . We define the coordinate transformations,*

$$\begin{aligned}
 \begin{pmatrix} W_n \\ Z_n \end{pmatrix} &= \begin{pmatrix} (wX_n + V_n) - (wY_n + U_n) \\ (zX_n + V_n) - (zY_n + U_n) \end{pmatrix}, \\
 \begin{pmatrix} W_{n+1} \\ Z_{n+1} \end{pmatrix} &= \begin{pmatrix} (wX_{n+1} + V_{n+1}) - (wY_{n+1} + U_{n+1}) \\ (zX_{n+1} + V_{n+1}) - (zY_{n+1} + U_{n+1}) \end{pmatrix},
 \end{aligned} \tag{54}$$

where $w \in [0, \frac{1}{2}\gamma)$ and $z = \gamma - w$. Then, under the assumptions (A.1) and (A.2), we have

$$(\|W_{n+1}\|_{\mathbb{L}_2}^2 + \|Z_{n+1}\|_{\mathbb{L}_2}^2) \leq (1 - 2\alpha h_n)(\|W_n\|_{\mathbb{L}_2}^2 + \|Z_n\|_{\mathbb{L}_2}^2), \quad (55)$$

for all $0 < h_n \leq 1$ and $n \geq 0$ where

$$\alpha = \frac{1}{z-w} \min \left[z^2 - uM_1, um - w^2, (\lambda_- z^2 - \lambda_+ w^2) - \frac{1}{2}u(M_1 - m) \right] - Kh_n, \quad (56)$$

for some constant $K > 0$ not depending on h_n .

Proof From the decomposition (44) and coordinate transformations (49) and (52), we have

$$\begin{aligned} \begin{pmatrix} W_{n+1} \\ Z_{n+1} \end{pmatrix} &= \begin{pmatrix} (w\tilde{X}_{n+1} + \tilde{V}_{n+1}) - (w\tilde{Y}_{n+1} + \tilde{U}_{n+1}) \\ (z\tilde{X}_{n+1} + \tilde{V}_{n+1}) - (z\tilde{Y}_{n+1} + \tilde{U}_{n+1}) \end{pmatrix} + \begin{pmatrix} (w\bar{X}_n + \bar{V}_n) - (w\bar{Y}_n + \bar{U}_n) \\ (z\bar{X}_n + \bar{V}_n) - (z\bar{Y}_n + \bar{U}_n) \end{pmatrix} \\ &= \begin{pmatrix} W_n^{(2)} \\ Z_n^{(2)} \end{pmatrix} + \begin{pmatrix} W_n^{(3)} \\ Z_n^{(3)} \end{pmatrix}. \end{aligned}$$

By applying the bounds (46), (50) and (53) given by the previous theorems along with the inequality (43) in Theorem 25 with $c = h_n^2$, we obtain

$$\begin{aligned} &(\|W_{n+1}\|_{\mathbb{L}_2}^2 + \|Z_{n+1}\|_{\mathbb{L}_2}^2) \\ &\leq (\|W_n^{(2)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(2)}\|_{\mathbb{L}_2}^2) + (\|W_n^{(3)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(3)}\|_{\mathbb{L}_2}^2) + 2\mathbb{E}[\langle W_n^{(2)}, W_n^{(3)} \rangle + \langle Z_n^{(2)}, Z_n^{(3)} \rangle] \\ &\leq (1 + h_n^2)(\|W_n^{(2)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(2)}\|_{\mathbb{L}_2}^2) + \left(1 + \frac{1}{h_n^2}\right) (\|W_n^{(3)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(3)}\|_{\mathbb{L}_2}^2) \\ &\leq (1 + h_n^2)(1 - \alpha_2 h_n)(\|W_n^{(1)}\|_{\mathbb{L}_2}^2 + \|Z_n^{(1)}\|_{\mathbb{L}_2}^2) + (h_n^2 + h_n^4)K_3(\|W_n\|_{\mathbb{L}_2}^2 + \|Z_n\|_{\mathbb{L}_2}^2) \\ &\leq ((1 + h_n^2)(1 - \alpha_1 h_n)(1 - \alpha_2 h_n) + (h_n^2 + h_n^4)K_3)(\|W_n\|_{\mathbb{L}_2}^2 + \|Z_n\|_{\mathbb{L}_2}^2) \\ &\leq (1 - 2\alpha h_n)(\|W_n\|_{\mathbb{L}_2}^2 + \|Z_n\|_{\mathbb{L}_2}^2), \end{aligned}$$

where

$$\begin{aligned} \alpha &= \frac{(\lambda_+ z^2 - \frac{1}{2}uM_1) \wedge (\frac{1}{2}um - \lambda_+ w^2)}{z-w} + \frac{(\lambda_- z^2 - \frac{1}{2}uM_1) \wedge (\frac{1}{2}um - \lambda_- w^2)}{z-w} - Kh_n \\ &= \frac{1}{z-w} \min \left[z^2 - uM_1, \lambda_+ z^2 - \lambda_- w^2 - \frac{1}{2}uM_1 + \frac{1}{2}um, \right. \\ &\quad \left. \lambda_- z^2 - \lambda_+ w^2 - \frac{1}{2}uM_1 + \frac{1}{2}um, um - w^2 \right] - Kh_n \\ &= \frac{1}{z-w} \min \left[z^2 - uM_1, um - w^2, (\lambda_- z^2 - \lambda_+ w^2) - \frac{1}{2}u(M_1 - m) \right] - Kh_n, \end{aligned}$$

and

$$K = 1 + \alpha_1 + \alpha_2 + \alpha_1 \cdot \alpha_2 + K_1 + K_2 + K_3. \quad \blacksquare$$

Remark 30 *Provided that $\gamma^2 > \frac{3+\sqrt{3}}{2}uM_1$, the contraction rate α becomes positive if $w = 0$ and h_n is sufficiently small.*

Remark 31 *We can recover the result from Theorem 8 using the inequality $1 - x \leq e^{-x}$. Although it is not strictly necessary for the proof of the global error, it does establish how the QUICSORT method undergoes exponential contractivity at each time interval $[t_n, t_{n+1}]$.*

Appendix C. Local error bounds

In this section we prove the local error bounds in Theorem 10. To do so, we compare the Taylor expansions of underdamped Langevin dynamics (2) and the QUICSORT scheme (3).

Theorem 32 (Stochastic Taylor expansion of Langevin dynamics) *Let $\{x_t, v_t\}_{t \geq 0}$ be the underdamped Langevin diffusion (2). Then, for all $n \geq 0$, over any interval $[t_n, t_{n+1}]$,*

$$x_{t_{n+1}} = x_{t_n} + h_n v_{t_n} + \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + R_x^{(1)}, \quad (57a)$$

$$\begin{aligned} &= x_{t_n} + h_n v_{t_n} + \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 - \frac{1}{2} h_n^2 (-\gamma v_{t_n} - u \nabla f(x_{t_n})) \\ &\quad - \gamma \sigma \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + R_x^{(2)}, \end{aligned} \quad (57b)$$

$$\begin{aligned} &= x_{t_n} + h_n v_{t_n} + \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + \frac{1}{2} h_n^2 (-\gamma v_{t_n} - u \nabla f(x_{t_n})) \\ &\quad - \gamma \sigma \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + \frac{1}{6} h_n^3 (\gamma^2 v_{t_n} + \gamma u \nabla f(x_{t_n}) - u \nabla^2 f(x_{t_n}) v_{t_n}) \\ &\quad + \sigma (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} W_{t_n, r_3} dr_3 dr_2 dr_1 + R_x^{(3)}. \end{aligned} \quad (57c)$$

$$v_{t_{n+1}} = v_{t_n} + \sigma W_n + h_n (-\gamma v_{t_n} - u \nabla f(x_{t_n})) - \gamma \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + R_v^{(1)}, \quad (58a)$$

$$\begin{aligned} &= v_{t_n} + \sigma W_n + h_n (-\gamma v_{t_n} - u \nabla f(x_{t_n})) - \gamma \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 \\ &\quad + \frac{1}{2} h_n^2 (\gamma^2 v_{t_n} + \gamma u \nabla f(x_{t_n}) - u \nabla^2 f(x_{t_n}) v_{t_n}) \\ &\quad + \sigma (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + R_v^{(2)}, \end{aligned} \quad (58b)$$

$$\begin{aligned} &= v_{t_n} + \sigma W_n + h_n (-\gamma v_{t_n} - u \nabla f(x_{t_n})) - \gamma \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1, \\ &\quad + \frac{1}{2} h_n^2 (\gamma^2 v_{t_n} + \gamma u \nabla f(x_{t_n}) - u \nabla^2 f(x_{t_n}) v_{t_n}) \\ &\quad + \sigma (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + \frac{1}{6} h_n^3 (-\gamma^3 v_{t_n} - \gamma^2 u \nabla f(x_{t_n})) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{6} h_n^3 \left(2\gamma u \nabla^2 f(x_{t_n}) v_{t_n} - u^2 \nabla f(x_{t_n}) \nabla f(x_{t_n}) - u \nabla^3 f(x_{t_n}) (v_{t_n}, v_{t_n}) \right) \\
 & - \gamma \sigma (\gamma^2 - 2u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} W_{t_n, r_3} dr_3 dr_2 dr_1 \\
 & - \sigma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^3 f(x_{t_n}) (v_{r_2}, W_{t_n, r_3}) dr_3 dr_2 dr_1 \\
 & - \sigma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \nabla^3 f(x_{t_n}) (v_{t_n}, (r_2 - t_n) W_{t_n, r_2}) dr_2 dr_1 + R_v^{(3)}, \tag{58c}
 \end{aligned}$$

where the remainder terms $R_x^{(1)}$, $R_x^{(2)}$ and $R_x^{(3)}$ for the position component are given by

$$R_x^{(1)} := -\gamma \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} v_{r_2} dr_2 dr_1 - u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \nabla f(x_{r_2}) dr_2 dr_1, \tag{59a}$$

$$\begin{aligned}
 R_x^{(2)} & := \gamma^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} v_{r_3} dr_3 dr_2 dr_1 + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{t_{n+1}} \nabla f(x_{r_3}) dr_3 dr_2 dr_1 \\
 & - u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^2 f(x_{r_3}) v_{r_3} dr_3 dr_2 dr_1, \tag{59b}
 \end{aligned}$$

$$\begin{aligned}
 R_x^{(3)} & := -\gamma (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} v_{r_4} dr_4 dr_3 dr_2 dr_1 \\
 & - u (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla f(x_{r_4}) dr_4 dr_3 dr_2 dr_1 \\
 & + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla^2 f(x_{r_4}) v_{r_4} dr_4 dr_3 dr_2 dr_1 \\
 & - u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla^3 f(x_{r_4}) (v_{r_4}, v_{r_3}) dr_4 dr_3 dr_2 dr_1, \tag{59c}
 \end{aligned}$$

and the remainder terms $R_v^{(1)}$, $R_v^{(2)}$ and $R_v^{(3)}$ for the momentum component are given by

$$\begin{aligned}
 R_v^{(1)} & := \gamma^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} v_{r_2} dr_2 dr_1 + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \nabla f(x_{r_2}) dr_2 dr_1 \\
 & - u \int_{t_n}^{t_{n+1}} (\nabla f(x_{r_1}) - \nabla f(x_{t_n})) dr_1, \tag{60a}
 \end{aligned}$$

$$\begin{aligned}
 R_v^{(2)} & := -\gamma (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} v_{r_3} dr_3 dr_2 dr_1 \\
 & - u (\gamma^2 - u \nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla f(x_{r_3}) dr_3 dr_2 dr_1 \\
 & + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^2 f(x_{r_3}) v_{r_3} dr_3 dr_2 dr_1 \\
 & - u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} (\nabla^2 f(x_{r_2}) v_{r_2} - \nabla^2 f(x_{t_n}) v_{r_2}) dr_2 dr_1, \tag{60b}
 \end{aligned}$$

$$\begin{aligned}
 R_v^{(3)} &:= \gamma^2(\gamma^2 - 2u\nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} v_{r_4} dr_4 dr_3 dr_2 dr_1 \\
 &+ \gamma u(\gamma^2 - 2u\nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla f(x_{r_4}) dr_4 dr_3 dr_2 dr_1 \\
 &- u(\gamma^2 - u\nabla^2 f(x_{t_n})) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla^2 f(x_{r_4}) v_{r_4} dr_4 dr_3 dr_2 dr_1 \\
 &+ \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla^3 f(x_{r_4})(v_{r_4}, v_{r_3}) dr_4 dr_3 dr_2 dr_1 \\
 &+ \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla^3 f(x_{t_n})(v_{r_4}, v_{r_2}) dr_4 dr_3 dr_2 dr_1 \\
 &+ u^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \nabla^3 f(x_{t_n})(\nabla f(x_{r_4}), v_{r_2}) dr_4 dr_3 dr_2 dr_1 \\
 &+ \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^3 f(x_{t_n})(v_{r_3}, (r_2 - t_n)v_{t_n}) dr_3 dr_2 dr_1 \\
 &+ u^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^3 f(x_{t_n})(\nabla f(x_{r_3}), (r_2 - t_n)v_{t_n}) dr_3 dr_2 dr_1 \\
 &- u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} (\nabla^3 f(x_{r_3})(v_{r_3}, v_{r_2}) - \nabla^3 f(x_{t_n})(v_{r_3}, v_{r_2})) dr_3 dr_2 dr_1. \quad (60c)
 \end{aligned}$$

Proof For any $r \in [t_n, t_{n+1}]$, we can rewrite the SDE (2) into integral form:

$$\begin{aligned}
 x_r &= x_{t_n} + \int_{t_n}^r v_{r_1} dr_1, \\
 v_r &= v_{t_n} - \gamma \int_{t_n}^r v_{r_1} dr_1 - u \int_{t_n}^r \nabla f(x_{r_1}) dr_1 + \sigma W_{t_n, r}.
 \end{aligned}$$

Assuming ∇f is sufficiently differentiable, by the fundamental theorem of calculus, we have

$$\begin{aligned}
 \nabla f(x_{r_i}) &= \nabla f(x_{t_n}) + \int_{t_n}^{r_i} \nabla^2 f(x_{r_{i+1}}) dx_{r_{i+1}} \\
 &= \nabla f(x_{t_n}) + \int_{t_n}^{r_i} \nabla^2 f(x_{r_{i+1}}) v_{r_{i+1}} dr_{i+1}, \\
 \nabla^2 f(x_{r_i}) v_r &= \nabla^2 f(x_{t_n}) v_r + \int_{t_n}^{r_i} \nabla^3 f(x_{r_{i+1}})(v_r, dx_{r_{i+1}}) \\
 &= \nabla^2 f(x_{t_n}) v_r + \int_{t_n}^{r_i} \nabla^3 f(x_{r_{i+1}})(v_r, v_{r_{i+1}}) dr_{i+1},
 \end{aligned}$$

for $t_n \leq r_4 \leq r_3 \leq r_2 \leq r_1 \leq t_{n+1}$. The theorem now follows by repeatedly substituting the above identities into the integral form of the SDE and rearranging the resulting terms. \blacksquare

Theorem 33 (ULD Remainder Bounds) *Let $\{(x_t, v_t)\}_{t \geq 0}$ be the underdamped Langevin diffusion (2) where we assume that $(x_0, v_0) \sim \pi$, the stationary distribution of the process. Let $0 < h_n \leq 1$ be fixed and consider the remainder terms given by (59a) and (60a), defined over the interval $[t_n, t_{n+1}]$. Under assumptions (A.1) and (A.2), we can bound their \mathbb{L}_2 norms and absolute means as*

$$\|R_x^{(1)}\|_{\mathbb{L}_2} \leq C_x^{(1)} \sqrt{d} h_n^2, \quad \|R_v^{(1)}\|_{\mathbb{L}_2} \leq C_v^{(1)} \sqrt{d} h_n^2, \quad (61)$$

$$\|\mathbb{E}[R_x^{(1)}]\|_2 \leq C_x^{(1)} \sqrt{d} h_n^2, \quad \|\mathbb{E}[R_v^{(1)}]\|_2 \leq C_v^{(1)} \sqrt{d} h_n^2, \quad (62)$$

where $C_x^{(1)}, C_v^{(1)} > 0$ are constants depending on γ, u, m and M_1 .

Under the additional assumption (A.3), we can bound the terms (59b) and (60b) as

$$\|R_x^{(2)}\|_{\mathbb{L}_2} \leq C_x^{(2)} \sqrt{d} h_n^3, \quad \|R_v^{(2)}\|_{\mathbb{L}_2} \leq C_v^{(2)} d h_n^3, \quad (63)$$

$$\|\mathbb{E}[R_x^{(2)}]\|_2 \leq C_x^{(2)} \sqrt{d} h_n^3, \quad \|\mathbb{E}[R_v^{(2)}]\|_2 \leq C_v^{(2)} d h_n^3, \quad (64)$$

where $C_x^{(2)}, C_v^{(2)} > 0$ are constants which depend on γ, u, m, M_1 and M_2 . Finally, under the additional assumption (A.4), the remainder terms (59c) and (60c) can be bounded by

$$\|R_x^{(3)}\|_{\mathbb{L}_2} \leq C_x^{(3)} d h_n^4, \quad \|R_v^{(3)}\|_{\mathbb{L}_2} \leq C_v^{(3)} d^{1.5} h_n^4, \quad (65)$$

$$\|\mathbb{E}[R_x^{(3)}]\|_2 \leq C_x^{(3)} d h_n^4, \quad \|\mathbb{E}[R_v^{(3)}]\|_2 \leq C_v^{(3)} d^{1.5} h_n^4, \quad (66)$$

where $C_x^{(3)}, C_v^{(3)} > 0$ are constants depending on γ, u, m, M_1, M_2 and M_3 .

Proof Suppose that $\gamma^2 \geq 2uM_1$. Then for all $x, v \in \mathbb{R}^d$,

$$\|(\gamma^2 - 2u\nabla^2 f(x))v\|_{\mathbb{L}_2} \leq \|(\gamma^2 - u\nabla^2 f(x))v\|_{\mathbb{L}_2} \leq (\gamma^2 - um)\|v\|_{\mathbb{L}_2}.$$

We can derive the local error bounds (61), (63) and (65) for the position variable using this inequality, Fubini's theorem and the bounds (28), (31), (37) and (40) given in Appendix A.

$$\begin{aligned} \|R_x^{(1)}\|_{\mathbb{L}_2} &\leq \gamma \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \|v_{r_2}\|_{\mathbb{L}_2} dr_2 dr_1 + u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \|\nabla f(x_{r_2})\|_{\mathbb{L}_2} dr_2 dr_1 \\ &\leq \frac{1}{2} h_n^2 (\gamma \sqrt{ud} + u \sqrt{M_1 d}) \\ &= C_x^{(1)} \sqrt{d} h_n^2, \end{aligned}$$

$$\begin{aligned} \|R_x^{(2)}\|_{\mathbb{L}_2} &\leq \gamma^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|v_{r_3}\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\ &\quad + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla f(x_{r_3})\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\ &\quad + u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla^2 f(x_{r_3})v_{r_3}\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{1}{6}h_n^3\left(\gamma^2\sqrt{ud} + \gamma u\sqrt{M_1d} + uM_1\sqrt{ud}\right) \\
 &= C_x^{(2)}\sqrt{d}h_n^3, \\
 \|R_x^{(3)}\|_{\mathbb{L}_2} &\leq \gamma(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|v_{r_4}\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + u(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla f(x_{r_4})\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla^2 f(x_{r_4})v_{r_4}\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla^3 f(x_{r_4})(v_{r_4}, v_{r_3})\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\leq \frac{1}{24}h_n^4\left(\gamma(\gamma^2 - um)\sqrt{ud} + u(\gamma^2 - um)\sqrt{M_1d} + \gamma uM_1\sqrt{ud} + \sqrt{3}M_2ud\right) \\
 &\leq \frac{1}{24}\left(\gamma(\gamma^2 - um)\sqrt{u} + u(\gamma^2 - um)\sqrt{M_1} + \gamma uM_1\sqrt{u} + \sqrt{3}M_2u\right)dh_n^4 \\
 &= C_x^{(3)}dh_n^4.
 \end{aligned}$$

For the momentum variable, we can additionally use (42) and the bound

$$\|x_t - x_{t_n}\|_{\mathbb{L}_{2p}} \leq \int_{t_n}^t \|v_{r_1}\|_{\mathbb{L}_{2p}} dr_1,$$

to show that

$$\begin{aligned}
 \|R_v^{(1)}\|_{\mathbb{L}_2} &\leq \gamma^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \|v_{r_2}\|_{\mathbb{L}_2} dr_2 dr_1 + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \|\nabla f(x_{r_2})\|_{\mathbb{L}_2} dr_2 dr_1 \\
 &\quad + u \int_{t_n}^{t_{n+1}} \|\nabla f(x_{r_1}) - \nabla f(x_{t_n})\|_{\mathbb{L}_2} dr_1 \\
 &\leq \frac{1}{2}h_n^2\left(\gamma^2\sqrt{ud} + \gamma u\sqrt{M_1d} + uM_1\sqrt{ud}\right) \\
 &= C_v^{(1)}\sqrt{d}h_n^2,
 \end{aligned}$$

$$\begin{aligned}
 \|R_v^{(2)}\|_{\mathbb{L}_2} &\leq \gamma(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|v_{r_3}\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\
 &\quad + u(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla f(x_{r_3})\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\
 &\quad + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla^2 f(x_{r_3})v_{r_3}\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\
 &\quad + u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \|\nabla^2 f(x_{r_2})v_{r_2} - \nabla^2 f(x_{t_n})v_{r_2}\|_{\mathbb{L}_2} dr_2 dr_1
 \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{1}{6} h_n^3 \left(\gamma(\gamma^2 - um) \sqrt{ud} + u(\gamma^2 - um) \sqrt{M_1 d} + \gamma u M_1 \sqrt{ud} + \sqrt{3} M_2 u d \right) \\
 &\leq \frac{1}{6} \left(\gamma(\gamma^2 - um) \sqrt{u} + u(\gamma^2 - um) \sqrt{M_1} + \gamma u M_1 \sqrt{u} + \sqrt{3} M_2 u \right) d h_n^3 \\
 &= C_v^{(2)} d h_n^3,
 \end{aligned}$$

$$\begin{aligned}
 \|R_v^{(3)}\|_{\mathbb{L}_2} &\leq \gamma^2(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_4} \|v_{r_4}\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + \gamma u(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_4} \|\nabla f(x_{r_4})\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + u(\gamma^2 - um) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla^2 f(x_{r_4}) v_{r_4}\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla^3 f(x_{r_4})(v_{r_4}, v_{r_3})\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla^3 f(x_{t_n})(v_{r_4}, v_{r_2})\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + u^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \int_{t_n}^{r_3} \|\nabla^3 f(x_{t_n})(\nabla f(x_{r_4}), v_{r_2})\|_{\mathbb{L}_2} dr_4 dr_3 dr_2 dr_1 \\
 &\quad + \gamma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla^3 f(x_{t_n})(v_{r_3}, (r_2 - t_n)v_{t_n})\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\
 &\quad + u^2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla^3 f(x_{t_n})(\nabla f(x_{r_3}), (r_2 - t_n)v_{t_n})\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\
 &\quad - u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|\nabla^3 f(x_{r_3})(v_{r_3}, v_{r_2}) - \nabla^3 f(x_{t_n})(v_{r_3}, v_{r_2})\|_{\mathbb{L}_2} dr_3 dr_2 dr_1 \\
 &\leq \frac{1}{24} h_n^4 \left(\gamma^2(\gamma^2 - um) \sqrt{ud} + \gamma u(\gamma^2 - um) \sqrt{M_1 d} + u(\gamma^2 - um) M_1 \sqrt{ud} \right. \\
 &\quad \left. + 3\sqrt{3} \gamma u^2 M_2 d + 2\sqrt{3} u^{2.5} \sqrt{M_1} M_2 d + \sqrt{15} u^{1.5} M_3 d^{1.5} \right) \\
 &\leq \frac{1}{24} \left(\gamma^2(\gamma^2 - um) \sqrt{u} + \gamma u^{1.5}(\gamma^2 - um) \sqrt{M_1} + u(\gamma^2 - um) M_1 \sqrt{u} \right. \\
 &\quad \left. + 3\sqrt{3} \gamma u^2 M_2 + 2\sqrt{3} u^{2.5} \sqrt{M_1} M_2 + \sqrt{15} u^{1.5} M_3 \right) d^{1.5} h_n^4 \\
 &= C_v^{(3)} d^{1.5} h_n^4.
 \end{aligned}$$

In order to prove the bounds (62), (64) and (66) for the mean errors, we simply note that

$$\|\mathbb{E}[R]\|_2 \leq \mathbb{E}[\|R\|_2] \leq \mathbb{E}[\|R\|_2^2]^{\frac{1}{2}} = \|R\|_{\mathbb{L}_2},$$

by Jensen's inequality. ■

Theorem 34 (Taylor expansion of the QUICSORT method) *Let $\{X_n, V_n\}_{n \geq 0}$ denote the QUICSORT method (3). Then for all $n \geq 0$, we have the following Taylor expansions:*

$$X_{n+1} = X_n + h_n V_n + \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + R_X^{(1)}, \quad (67a)$$

$$\begin{aligned} &= X_n + h_n V_n + \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + \frac{1}{2} h_n^2 [-\gamma V_n - u \nabla f(X_n)] \\ &\quad - \gamma \sigma \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + R_X^{(2)}, \end{aligned} \quad (67b)$$

$$\begin{aligned} &= X_n + h_n V_n + \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + \frac{1}{2} h_n^2 [-\gamma V_n - u \nabla f(X_n)] \\ &\quad - \gamma \sigma \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + \frac{1}{6} h_n^3 [\gamma^2 V_n + \gamma u \nabla f(X_n) - u \nabla^2 f(X_n) V_n] \\ &\quad + \sigma (\gamma^2 - u \nabla^2 f(X_n)) \left[\frac{1}{24} h_n^3 W_n + \frac{1}{6} h_n^3 H_n + \frac{1}{2} h_n^3 K_n \right] + R_X^{(3)}, \end{aligned} \quad (67c)$$

$$V_{n+1} = V_n + \sigma W_n + h_n [-\gamma V_n - u \nabla f(X_n)] - \gamma \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 + R_V^{(1)}, \quad (68a)$$

$$\begin{aligned} &= V_n + \sigma W_n + h_n [-\gamma V_n - u \nabla f(X_n)] - \gamma \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 \\ &\quad + \frac{1}{2} h_n^2 [\gamma^2 V_n + \gamma u \nabla f(X_n) - u \nabla^2 f(X_n) V_n] \\ &\quad + \sigma (\gamma^2 - u \nabla^2 f(X_n)) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 + R_V^{(2)}, \end{aligned} \quad (68b)$$

$$\begin{aligned} &= V_n + \sigma W_n + h_n [-\gamma V_n - u \nabla f(X_n)] - \gamma \sigma \int_{t_n}^{t_{n+1}} W_{t_n, r_1} dr_1 \\ &\quad + \frac{1}{2} h_n^2 [\gamma^2 V_n + \gamma u \nabla f(X_n) - u \nabla^2 f(X_n) V_n] \\ &\quad + \sigma (\gamma^2 - u \nabla^2 f(X_n)) \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} W_{t_n, r_2} dr_2 dr_1 - \frac{1}{6} h_n^3 [\gamma^3 V_n + \gamma^2 u \nabla f(X_n)] \\ &\quad + \frac{1}{6} h_n^3 [2\gamma u \nabla^2 f(X_n) V_n - u^2 \nabla f(X_n) \nabla f(X_n) - u \nabla^3 f(X_n) (V_n, V_n)] \\ &\quad - \gamma \sigma (\gamma^2 - 2u \nabla^2 f(X_n)) \left[\frac{1}{24} h_n^3 W_n + \frac{1}{6} h_n^3 H_n + \frac{1}{2} h_n^3 K_n \right] \\ &\quad - \sigma u \nabla^3 f(X_n) \left(V_n, \left[\frac{1}{8} h_n^3 W_n + \frac{1}{3} h_n^3 H_n + \frac{1}{2} h_n^3 K_n \right] \right) + R_V^{(3)}, \end{aligned} \quad (68c)$$

where

$$R_X^{(1)} = (\phi_1(1) - h_n)(V_n + \sigma(H_n + 6K_n)) + \left[\frac{1}{h_n} \phi_2(1) - \frac{1}{2} h_n \right] \sigma(W_n - 12K_n)$$

$$\begin{aligned}
 & -\frac{1}{2}uh_n(\phi_1(\lambda_+) + \phi_1(\lambda_-))\nabla f(X_n) \\
 & -\frac{1}{2}uh_n\left[\phi_1(\lambda_+)I_1^{(-)} + \phi_1(\lambda_-)I_1^{(+)} + \phi_1(\lambda_-)\tilde{I}_1\right], \tag{69a}
 \end{aligned}$$

$$\begin{aligned}
 R_X^{(2)} &= \left[\phi_1(1) - \left(h_n - \frac{1}{2}\gamma h_n^2\right)\right](V_n + \sigma(H_n + 6K_n)) \\
 &+ \left[\frac{1}{h_n}\phi_2(1) - \left(\frac{1}{2}h_n - \frac{1}{6}\gamma h_n^2\right)\right]\sigma(W_n - 12K_n) \\
 &+ \left[-\frac{1}{2}h_n(\phi_1(\lambda_+) + \phi_1(\lambda_-)) - \left(-\frac{1}{2}h_n^2\right)\right]u\nabla f(X_n) \\
 &- \frac{1}{2}uh_n\left[\phi_1(\lambda_+)I_1^{(-)} + \phi_1(\lambda_-)I_1^{(+)} + \phi_1(\lambda_-)\tilde{I}_1\right], \tag{69b}
 \end{aligned}$$

$$\begin{aligned}
 R_X^{(3)} &= \left[\phi_1(1) - \left(h_n - \frac{1}{2}\gamma h_n^2 + \frac{1}{6}\gamma^2 h_n^3\right)\right](V_n + \sigma(H_n + 6K_n)) \\
 &+ \left[\frac{1}{h_n}\phi_2(1) - \left(\frac{1}{2}h_n - \frac{1}{6}\gamma h_n^2 + \frac{1}{24}\gamma^2 h_n^3\right)\right]\sigma(W_n - 12K_n) \\
 &+ \left[-\frac{1}{2}h_n(\phi_1(\lambda_+) + \phi_1(\lambda_-)) - \left(-\frac{1}{2}h_n^2 + \frac{1}{6}\gamma h_n^3\right)\right]u\nabla f(X_n) \\
 &+ \left[-h_n\phi_1(\lambda_+)\phi_1(\lambda_-) - \left(-\frac{1}{6}h_n^3\right)\right]u\nabla^2 f(X_n)\left[V_n + \sigma(H_n + 6K_n)\right] \\
 &+ \left[-\frac{1}{2}(\phi_1(\lambda_+)\phi_2(\lambda_-) + \phi_1(\lambda_-)\phi_2(\lambda_+)) - \left(-\frac{1}{24}h_n^3\right)\right] \\
 &\quad \times \sigma u\nabla^2 f(X_n)(W_n - 12K_n) \\
 &- \frac{1}{2}uh_n\left[\phi_1(\lambda_+)I_2^{(-)} + \phi_1(\lambda_-)I_2^{(+)} + \phi_1(\lambda_-)\tilde{I}_1\right], \tag{69c}
 \end{aligned}$$

$$\begin{aligned}
 R_V^{(1)} &= (\phi_0(1) - (1 - \gamma h_n))(V_n + \sigma(H_n + 6K_n)) \\
 &+ \left[\frac{1}{h_n}\phi_1(1) - \left(1 - \frac{1}{2}\gamma h_n\right)\right]\sigma(W_n - 12K_n) \\
 &+ \left[-\frac{1}{2}h_n(\phi_0(\lambda_+) + \phi_0(\lambda_-)) - (-h_n)\right]u\nabla f(X_n) \\
 &- \frac{1}{2}uh_n\left[\phi_0(\lambda_+)I_1^{(-)} + \phi_0(\lambda_-)I_1^{(+)} + \phi_0(\lambda_-)\tilde{I}_1\right], \tag{69d}
 \end{aligned}$$

$$\begin{aligned}
 R_V^{(2)} &= \left[\phi_0(1) - (1 - \gamma h_n + \frac{1}{2}\gamma^2 h_n^2)\right](V_n + \sigma(H_n + 6K_n)) \\
 &+ \left[\frac{1}{h_n}\phi_1(1) - \left(1 - \frac{1}{2}\gamma h_n + \frac{1}{6}\gamma^2 h_n^2\right)\right]\sigma(W_n - 12K_n) \\
 &+ \left[-\frac{1}{2}h_n(\phi_0(\lambda_+) + \phi_0(\lambda_-)) - \left(-h_n + \frac{1}{2}\gamma h_n^2\right)\right]u\nabla f(X_n)
 \end{aligned}$$

$$\begin{aligned}
 & + \left[-\frac{1}{2}h_n(\phi_0(\lambda_+)\phi_1(\lambda_-) + \phi_0(\lambda_-)\phi_1(\lambda_+)) - \left(-\frac{1}{2}h_n^2\right) \right] \\
 & \quad \times u\nabla^2 f(X_n)(V_n + \sigma(H_n + 6K_n)) \\
 & + \left[-\frac{1}{2}(\phi_0(\lambda_+)\phi_2(\lambda_-) + \phi_0(\lambda_-)\phi_2(\lambda_+)) - \left(-\frac{1}{6}h_n^2\right) \right] \\
 & \quad \times \sigma u\nabla^2 f(X_n)(W_n - 12K_n) \\
 & - \frac{1}{2}uh_n \left[\phi_0(\lambda_+)I_2^{(-)} + \phi_0(\lambda_-)I_2^{(+)} + \phi_0(\lambda_-)\tilde{I}_1 \right], \tag{69e}
 \end{aligned}$$

$$\begin{aligned}
 R_V^{(3)} = & \left[\phi_0(1) - \left(1 - \gamma h_n + \frac{1}{2}\gamma^2 h_n^2 - \frac{1}{6}\gamma^3 h_n^3\right) \right] (V_n + \sigma(H_n + 6K_n)) \\
 & + \left[\frac{1}{h_n}\phi_1(1) - \left(1 - \frac{1}{2}\gamma h_n + \frac{1}{6}\gamma^2 h_n^2 - \frac{1}{24}\gamma^3 h_n^3\right) \right] \sigma(W_n - 12K_n) \\
 & + \left[-\frac{1}{2}h_n(\phi_0(\lambda_+) + \phi_0(\lambda_-)) - \left(-h + \frac{1}{2}\gamma h^2 - \frac{1}{6}\gamma^2 h^3\right) \right] u\nabla f(X_n) \\
 & + \left[-\frac{1}{2}h_n(\phi_0(\lambda_+)\phi_1(\lambda_-) + \phi_0(\lambda_-)\phi_1(\lambda_+)) - \left(-\frac{1}{2}h_n^2 + \frac{1}{3}\gamma h_n^3\right) \right] \\
 & \quad \times u\nabla^2 f(X_n)(V_n + \sigma(H_n + 6K_n)) \\
 & + \left[-\frac{1}{2}(\phi_0(\lambda_+)\phi_2(\lambda_-) + \phi_0(\lambda_-)\phi_2(\lambda_+)) - \left(-\frac{1}{6}h_n^2 + \frac{1}{12}\gamma h_n^3\right) \right] \\
 & \quad \times \sigma u\nabla^2 f(X_n)(W_n - 12K_n) \\
 & + \left[\frac{1}{2}h_n^2\phi_0(\lambda_-)\phi_1\left(\frac{1}{3}\right) - \frac{1}{6}h_n^3 \right] u^2\nabla^2 f(X_n)\nabla f(X_n) \\
 & + \left[-\frac{1}{4}h_n(\phi_0(\lambda_+)\phi_1^2(\lambda_-) + \phi_0(\lambda_-)\phi_1^2(\lambda_+)) - \left(-\frac{1}{6}h_n^3\right) \right] \\
 & \quad \times u\nabla^3 f(X_n)(V_n^{\otimes 2} + 2\sigma V_n \otimes (H_n + 6K_n)) \\
 & + \left[-\frac{1}{4}h_n(\phi_0(\lambda_+)\phi_1^2(\lambda_-) + \phi_0(\lambda_-)\phi_1^2(\lambda_+)) \right] \sigma^2 u\nabla^3 f(X_n)(H_n + 6K_n)^{\otimes 2} \\
 & + \left[-\frac{1}{2}(\phi_0(\lambda_+)\phi_1(\lambda_-)\phi_2(\lambda_-) + \phi_0(\lambda_-)\phi_1(\lambda_+)\phi_2(\lambda_+)) - \left(-\frac{1}{8}h_n^3\right) \right] \\
 & \quad \times \sigma u\nabla^3 f(X_n)(V_n, W_n - 12K_n) \\
 & + \left[-\frac{1}{2}(\phi_0(\lambda_+)\phi_1(\lambda_-)\phi_2(\lambda_-) + \phi_0(\lambda_-)\phi_1(\lambda_+)\phi_2(\lambda_+)) \right] \\
 & \quad \times \sigma^2 u\nabla^3 f(X_n)(H_n + 6K_n, W_n - 12K_n) \\
 & + \left[-\frac{1}{4} \cdot \frac{1}{h_n}(\phi_0(\lambda_+)\phi_2^2(\lambda_-) + \phi_0(\lambda_-)\phi_2^2(\lambda_+)) \right] \sigma^2 u\nabla^3 f(X_n)(W_n - 12K_n)^{\otimes 2} \\
 & - \frac{1}{2}uh_n \left[\phi_0(\lambda_+)I_3^{(-)} + \phi_0(\lambda_-)I_3^{(+)} + \phi_0(\lambda_-)\tilde{I}_2 + \phi_0(\lambda_-)\tilde{I}_3 \right]
 \end{aligned}$$

$$+ \frac{1}{2} u^2 h_n^2 \phi_0(\lambda_-) \phi_1\left(\frac{1}{3}\right) \nabla^2 f(X_n) \tilde{I}^{(1)}, \quad (69f)$$

for $\lambda_{\pm} = \frac{3 \pm \sqrt{3}}{6}$, along with the functions ϕ_0 , ϕ_1 and ϕ_2 defined in (35), and

$$\begin{aligned} I_1^{(\pm)} &:= \int_0^1 \nabla^2 f(X_n + tL_n^{(\pm)}) L_n^{(\pm)} dt \\ &= \left[\nabla f(X_n + L_n^{(\pm)}) - \nabla f(X_n) \right], \\ I_2^{(\pm)} &:= \int_0^1 (1-t) \nabla^3 f(X_n + tL_n^{(\pm)}) (L_n^{(\pm)})^{\otimes 2} dt \\ &= \int_0^1 \left[\nabla^2 f(X_n + L_n^{(\pm)}) - \nabla^2 f(X_n) \right] L_n^{(\pm)} dt, \\ I_3^{(\pm)} &:= \int_0^1 (1-t) \left[\nabla^3 f(X_n + tL_n^{(\pm)}) - \nabla^3 f(X_n) \right] (L_n^{(\pm)})^{\otimes 2} dt, \\ \tilde{I}_1 &:= \int_0^1 \nabla^2 f(X_n + L_n^{(+)} + tP_n) P_n dt \\ &= \left[\nabla f(X_n + L_n^{(+)} + P_n) - \nabla f(X_n + L_n^{(+)}) \right], \\ \tilde{I}_2 &:= \int_0^1 (1-t) \nabla^3 f(X_n + L_n^{(+)} + tP_n) (P_n)^{\otimes 2} dt \\ &= \int_0^1 \left[\nabla^2 f(X_n + L_n^{(+)} + tP_n) - \nabla^2 f(X_n + L_n^{(+)}) \right] P_n dt, \\ \tilde{I}_3 &:= \int_0^1 \nabla^3 f(X_n + tL_n^{(+)}) (L_n^{(+)}, P_n) dt \\ &= \left[\nabla^2 f(X_n + L_n^{(+)}) - \nabla^2 f(X_n) \right] P_n, \end{aligned} \quad (70)$$

where

$$\begin{aligned} L_n^{(\pm)} &:= \phi_1(\lambda_{\pm})(V_n + \sigma(H_n + 6K_n)) + \phi_2(\lambda_{\pm})C_n, \\ P_n &:= -uh_n \phi_1\left(\frac{1}{3}\right) \nabla f(X_n^{(1)}) = -uh_n \phi_1\left(\frac{1}{3}\right) \left(\nabla f(X_n) + I_1^{(-)} \right). \end{aligned}$$

Proof We can rewrite the QUICSORT method (3) as

$$\begin{aligned} X_{n+1} &= X_n + \phi_1(1)(V_n + \sigma(H_n + 6K_n)) + \phi_2(1)C_n \\ &\quad - \frac{1}{2} uh_n \left(\phi_1(\lambda_+) \nabla f(X_n^{(1)}) + \phi_1(\lambda_-) \nabla f(X_n^{(2)}) \right), \\ V_{n+1} &= \phi_0(1)(V_n + \sigma(H_n + 6K_n)) - \sigma(H_n - 6K_n) + \phi_1(1)C_n \\ &\quad - \frac{1}{2} uh_n \left(\phi_0(\lambda_+) \nabla f(X_n^{(1)}) + \phi_0(\lambda_-) \nabla f(X_n^{(2)}) \right), \end{aligned}$$

where

$$X_n^{(1)} = X_n + L_n^{(-)}, \quad X_n^{(2)} = X_n + L_n^{(+)} + P_n.$$

Using Taylor expansions with an integral remainder, we can derive the identities in (70),

$$\begin{aligned} \nabla f(X_n^{(1)}) &= \nabla f(X_n) + I_1^{(-)} \\ &= \nabla f(X_n) + \nabla^2 f(X_n) L_n^{(-)} + I_2^{(-)} \\ &= \nabla f(X_n) + \nabla^2 f(X_n) L_n^{(-)} + \frac{1}{2} \nabla^3 f(X_n) (L_n^{(-)}, L_n^{(-)}) + I_3^{(-)}, \\ \nabla f(X_n^{(2)}) &= \nabla f(X_n) + I_1^{(+)} + \tilde{I}_1 \\ &= \nabla f(X_n) + \nabla^2 f(X_n) L_n^{(+)} + I_2^{(+)} + \tilde{I}_1 \\ &= \nabla f(X_n) + \nabla^2 f(X_n) L_n^{(+)} + \nabla^2 f(X_n) P_n + \frac{1}{2} \nabla^3 f(X_n) (L_n^{(+)}, L_n^{(+)}) \\ &\quad + I_3^{(+)} + \tilde{I}_2 + \tilde{I}_3. \end{aligned}$$

The result now follows by substituting the expansions for $\nabla f(X_n^{(1)})$ and $\nabla f(X_n^{(2)})$ into the QUICSORT method, and Taylor expanding the ϕ functions (35) up to the desired order. ■

Theorem 35 (Remainder bounds for the QUICSORT method) *Let $\{(X_n, V_n)\}_{n \geq 0}$ denote the QUICSORT method (3) and let $\{(x_t, v_t)\}_{t \geq 0}$ denote the underdamped Langevin diffusion (2) where we assume that $(x_0, v_0) \sim \pi$, the stationary distribution of the process. Suppose that at time t_n , we have $(X_n, V_n) = (x_{t_n}, v_{t_n})$. Let $0 < h_n \leq 1$ be fixed, and consider the remainder terms (69a) and (69d) defined over the time interval $[t_n, t_{n+1}]$. Under assumptions (A.1) and (A.2) we can bound their \mathbb{L}_2 norms and absolute means as*

$$\|R_X^{(1)}\|_{\mathbb{L}_2} \leq C_X^{(1)} \sqrt{d} h_n^2, \quad \|R_V^{(1)}\|_{\mathbb{L}_2} \leq C_V^{(1)} \sqrt{d} h_n^2, \quad (71)$$

$$\|\mathbb{E}[R_X^{(1)}]\|_2 \leq C_X^{(1)} \sqrt{d} h_n^2, \quad \|\mathbb{E}[R_V^{(1)}]\|_2 \leq C_V^{(1)} \sqrt{d} h_n^2, \quad (72)$$

where $C_X^{(1)}, C_V^{(1)} > 0$ are constants depending on γ, u, m and M_1 .

Under the additional assumption (A.3), we can bound (69b) and (69e) as

$$\|R_X^{(2)}\|_{\mathbb{L}_2} \leq C_X^{(2)} \sqrt{d} h_n^3, \quad \|R_V^{(2)}\|_{\mathbb{L}_2} \leq C_V^{(2)} d h_n^3, \quad (73)$$

$$\|\mathbb{E}[R_X^{(2)}]\|_2 \leq C_X^{(2)} \sqrt{d} h_n^3, \quad \|\mathbb{E}[R_V^{(2)}]\|_2 \leq C_V^{(2)} d h_n^3, \quad (74)$$

where $C_X^{(2)}, C_V^{(2)} > 0$ are constants depending on γ, u, m, M_1 and M_2 . Finally, under the additional assumption (A.4), then the remainder terms (69c) and (69f) can be bounded by

$$\|R_X^{(3)}\|_{\mathbb{L}_2} \leq C_X^{(3)} d h_n^4, \quad \|R_V^{(3)}\|_{\mathbb{L}_2} \leq C_V^{(3)} d^{1.5} h_n^4, \quad (75)$$

$$\|\mathbb{E}[R_X^{(3)}]\|_2 \leq C_X^{(3)} d h_n^4, \quad \|\mathbb{E}[R_V^{(3)}]\|_2 \leq C_V^{(3)} d^{1.5} h_n^4, \quad (76)$$

where $C_X^{(3)}, C_V^{(3)} > 0$ are constants depending on γ, u, m, M_1, M_2 and M_3 .

Proof We can bound the integrals given in (70) using the Cauchy-Schwarz and Minkowski inequalities along with assumptions (A.1), (A.2), (A.3) and (A.4).

$$\begin{aligned}
 \|I_1^{(\pm)}\|_{\mathbb{L}_2} &\leq M_1 \|L_n^{(\pm)}\|_{\mathbb{L}_2} \\
 &\leq M_1 |\phi_1(\lambda_{\pm})| \|V_n\|_{\mathbb{L}_2} + \sigma M_1 |\phi_1(\lambda_{\pm})| \|H_n + 6K_n\|_{\mathbb{L}_2} \\
 &\quad + \sigma M_1 \left| \frac{1}{h_n} \phi_2(\lambda_{\pm}) \right| \|W_n - 12K_n\|_{\mathbb{L}_2}, \\
 \|I_2^{(\pm)}\|_{\mathbb{L}_2} &\leq M_2 \|L_n^{(\pm)}\|_{\mathbb{L}_4}^2 \\
 &\leq M_2 |\phi_1(\lambda_{\pm})|^2 \|V_n\|_{\mathbb{L}_4}^2 + \sigma^2 M_2 |\phi_1(\lambda_{\pm})|^2 \|H_n + 6K_n\|_{\mathbb{L}_4}^2 \\
 &\quad + \sigma^2 M_2 \left| \frac{1}{h_n} \phi_2(\lambda_{\pm}) \right|^2 \|W_n - 12K_n\|_{\mathbb{L}_4}^2 \\
 &\quad + 2\sigma M_2 |\phi_1(\lambda_{\pm})|^2 \|V_n\|_{\mathbb{L}_4} \|H_n + 6K_n\|_{\mathbb{L}_4} \\
 &\quad + 2\sigma M_2 \left| \frac{1}{h_n} \phi_1(\lambda_{\pm}) \phi_2(\lambda_{\pm}) \right| \|V_n\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4} \\
 &\quad + 2\sigma^2 M_2 \left| \frac{1}{h_n} \phi_1(\lambda_{\pm}) \phi_2(\lambda_{\pm}) \right| \|H_n + 6K_n\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4}, \\
 \|I_3^{(\pm)}\|_{\mathbb{L}_2} &\leq M_3 \|L_n^{(\pm)}\|_{\mathbb{L}_6}^3 \\
 &\leq M_3 |\phi_1(\lambda_{\pm})|^3 \|V_n\|_{\mathbb{L}_6}^3 + \sigma^3 M_3 |\phi_1(\lambda_{\pm})|^3 \|H_n + 6K_n\|_{\mathbb{L}_6}^3 \\
 &\quad + \sigma^3 M_3 \left| \frac{1}{h_n} \phi_2(\lambda_{\pm}) \right|^3 \|W_n - 12K_n\|_{\mathbb{L}_6}^3 \\
 &\quad + 3\sigma M_3 |\phi_1(\lambda_{\pm})|^3 \|V_n\|_{\mathbb{L}_6}^2 \|H_n + 6K_n\|_{\mathbb{L}_6} \\
 &\quad + 3\sigma^2 M_3 |\phi_1(\lambda_{\pm})|^3 \|V_n\|_{\mathbb{L}_6} \|H_n + 6K_n\|_{\mathbb{L}_6}^2 \\
 &\quad + 3\sigma M_3 \left| \frac{1}{h_n} \phi_1^2(\lambda_{\pm}) \phi_2(\lambda_{\pm}) \right| \|V_n\|_{\mathbb{L}_6}^2 \|W_n - 12K_n\|_{\mathbb{L}_6} \\
 &\quad + 3\sigma^3 M_3 \left| \frac{1}{h_n} \phi_1^2(\lambda_{\pm}) \phi_2(\lambda_{\pm}) \right| \|H_n + 6K_n\|_{\mathbb{L}_6}^2 \|W_n - 12K_n\|_{\mathbb{L}_6} \\
 &\quad + 3\sigma^2 M_3 \left| \frac{1}{h_n^2} \phi_1(\lambda_{\pm}) \phi_2^2(\lambda_{\pm}) \right| \|V_n\|_{\mathbb{L}_6} \|W_n - 12K_n\|_{\mathbb{L}_6}^2 \\
 &\quad + 3\sigma^3 M_3 \left| \frac{1}{h_n^2} \phi_1(\lambda_{\pm}) \phi_2^2(\lambda_{\pm}) \right| \|H_n + 6K_n\|_{\mathbb{L}_6} \|W_n - 12K_n\|_{\mathbb{L}_6}^2 \\
 &\quad + 6\sigma^2 M_3 \left| \frac{1}{h_n} \phi_1^2(\lambda_{\pm}) \phi_2(\lambda_{\pm}) \right| \|V_n\|_{\mathbb{L}_6} \|H_n + 6K_n\|_{\mathbb{L}_6} \|W_n - 12K_n\|_{\mathbb{L}_6}, \\
 \|\tilde{I}_1\|_{\mathbb{L}_2} &\leq M_1 \|P_n\|_{\mathbb{L}_2} \\
 &\leq uM_1 h_n \left| \phi_1\left(\frac{1}{3}\right) \right| \|\nabla f(X_n)\|_{\mathbb{L}_2} + uM_1^2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_-) \right| \|V_n\|_{\mathbb{L}_2} \\
 &\quad + \sigma uM_1^2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_-) \right| \|H_n + 6K_n\|_{\mathbb{L}_2} \\
 &\quad + \sigma uM_1^2 \left| \phi_1\left(\frac{1}{3}\right) \phi_2(\lambda_-) \right| \|W_n - 12K_n\|_{\mathbb{L}_2},
 \end{aligned}$$

$$\begin{aligned}
 \|\tilde{I}_2\|_{\mathbb{L}_2} &\leq M_2 \|P_n\|_{\mathbb{L}_4}^2 \\
 &\leq u^2 M_2 h_n^2 \left| \phi_1\left(\frac{1}{3}\right) \right|^2 \|\nabla f(X_n)\|_{\mathbb{L}_4}^2 \\
 &\quad + 2u^2 M_1 M_2 h_n^2 \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1(\lambda_-) \right| \|\nabla f(X_n)\|_{\mathbb{L}_4} \|V_n\|_{\mathbb{L}_4} \\
 &\quad + 2\sigma u^2 M_1 M_2 h_n^2 \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1(\lambda_-) \right| \|\nabla f(X_n)\|_{\mathbb{L}_4} \|H_n + 6K_n\|_{\mathbb{L}_4} \\
 &\quad + 2\sigma u^2 M_1 M_2 h_n \left| \phi_1^2\left(\frac{1}{3}\right) \phi_2(\lambda_-) \right| \|\nabla f(X_n)\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4} \\
 &\quad + u^2 M_1^2 M_2 h_n^2 \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1^2(\lambda_-) \right| \|V_n\|_{\mathbb{L}_4}^2 \\
 &\quad + \sigma^2 u^2 M_1^2 M_2 h_n^2 \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1^2(\lambda_-) \right| \|H_n + 6K_n\|_{\mathbb{L}_4}^2 \\
 &\quad + \sigma^2 u^2 M_1^2 M_2 \left| \phi_1^2\left(\frac{1}{3}\right) \phi_2^2(\lambda_-) \right| \|W_n - 12K_n\|_{\mathbb{L}_4}^2 \\
 &\quad + 2\sigma u^2 M_1^2 M_2 h_n^2 \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1^2(\lambda_-) \right| \|V_n\|_{\mathbb{L}_4} \|H_n + 6K_n\|_{\mathbb{L}_4} \\
 &\quad + 2\sigma u^2 M_1^2 M_2 h_n \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1(\lambda_-) \phi_2(\lambda_-) \right| \|V_n\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4} \\
 &\quad + 2\sigma^2 u^2 M_1^2 M_2 h_n \left| \phi_1^2\left(\frac{1}{3}\right) \phi_1(\lambda_-) \phi_2(\lambda_-) \right| \|H_n + 6K_n\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4},
 \end{aligned}$$

$$\begin{aligned}
 \|\tilde{I}_3\|_{\mathbb{L}_2} &\leq M_2 \|L_n^{(+)}\|_{\mathbb{L}_4} \|P_n\|_{\mathbb{L}_4} \\
 &\leq u M_2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_+) \right| \|\nabla f(X_n)\|_{\mathbb{L}_4} \|V_n\|_{\mathbb{L}_4} \\
 &\quad + \sigma u M_2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_+) \right| \|\nabla f(X_n)\|_{\mathbb{L}_4} \|H_n + 6K_n\|_{\mathbb{L}_4} \\
 &\quad + \sigma u M_2 \left| \phi_1\left(\frac{1}{3}\right) \phi_2(\lambda_+) \right| \|\nabla f(X_n)\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4} \\
 &\quad + u M_1 M_2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_+) \phi_1(\lambda_-) \right| \|V_n\|_{\mathbb{L}_4}^2 \\
 &\quad + \sigma^2 u M_1 M_2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_+) \phi_1(\lambda_-) \right| \|H_n + 6K_n\|_{\mathbb{L}_4}^2 \\
 &\quad + \sigma^2 u M_1 M_2 \left| \frac{1}{h_n} \phi_1\left(\frac{1}{3}\right) \phi_2(\lambda_+) \phi_2(\lambda_-) \right| \|W_n - 12K_n\|_{\mathbb{L}_4}^2 \\
 &\quad + 2\sigma u M_1 M_2 h_n \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_+) \phi_1(\lambda_-) \right| \|V_n\|_{\mathbb{L}_4} \|H_n + 6K_n\|_{\mathbb{L}_4} \\
 &\quad + \sigma u M_1 M_2 \left(\left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_+) \phi_2(\lambda_-) \right| + \left| \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_-) \phi_2(\lambda_+) \right| \right) \\
 &\quad \times \left[\|V_n\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4} + \sigma \|H_n + 6K_n\|_{\mathbb{L}_4} \|W_n - 12K_n\|_{\mathbb{L}_4} \right].
 \end{aligned}$$

We can derive the remainder bounds (71), (73) and (75) for the position variable using the above inequalities along with equations (28), (31) and (34). In addition, we will use Theorem 23 and Taylor expand the ϕ_i functions using the Lagrange form of the remainder.

$$\begin{aligned}
 \|R_X^{(1)}\|_{\mathbb{L}_2} &\leq \frac{1}{2}\gamma h_n^2 \|V_n\|_{\mathbb{L}_2} + \frac{1}{2}\gamma\sigma h_n^2 \|H_n + 6K_n\|_{\mathbb{L}_2} + \frac{1}{6}\gamma\sigma h_n^2 \|W_n - 12K_n\|_{\mathbb{L}_2} \\
 &\quad + \frac{1}{2}uh_n^2 \|\nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{2}uh_n \left(|\phi_1(\lambda_+)| \|I_1^{(-)}\|_{\mathbb{L}_2} + |\phi_1(\lambda_-)| \|I_1^{(+)}\|_{\mathbb{L}_2} \right) \\
 &\quad + \frac{1}{2}uh_n |\phi_1(\lambda_-)| \|\tilde{I}_1\|_{\mathbb{L}_2} \\
 &\leq \left[\frac{1}{2}\gamma\sqrt{u} + \frac{1}{2}u\sqrt{M_1} + \frac{2 + \sqrt{3} + \sqrt{5}}{12}\gamma\sigma\sqrt{h_n} + \frac{1}{6}u^{1.5}M_1h_n \right. \\
 &\quad + \frac{15 + 10\sqrt{3} + 9\sqrt{5}}{360}\sigma u M_1 h_n^{1.5} + \frac{3 - \sqrt{3}}{36}u^2 M_1^{1.5} h_n^2 \\
 &\quad \left. + \frac{2 - \sqrt{3}}{36}u^{2.5} M_1^2 h_n^3 + \frac{15 - 5\sqrt{3} + 21\sqrt{5} - 11\sqrt{15}}{2160}\sigma u^2 M_1^2 h_n^{3.5} \right] \sqrt{dh_n^2} \\
 &\leq \left[\gamma\sqrt{u} + \gamma\sigma + u\sqrt{M_1} + u^{1.5}M_1 + \sigma u M_1 + u^2 M_1^{1.5} + u^{2.5}M_1^2 + \sigma u^2 M_1^2 \right] \sqrt{dh_n^2}, \\
 \|R_X^{(2)}\|_{\mathbb{L}_2} &\leq \frac{1}{6}\gamma^2 h_n^3 \|V_n\|_{\mathbb{L}_2} + \frac{1}{6}\gamma^2\sigma h_n^3 \|H_n + 6K_n\|_{\mathbb{L}_2} + \frac{1}{24}\gamma^2\sigma h_n^3 \|W_n - 12K_n\|_{\mathbb{L}_2} \\
 &\quad + \frac{1}{6}\gamma u h_n^3 \|\nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{2}uh_n \left(|\phi_1(\lambda_+)| \|I_1^{(-)}\|_{\mathbb{L}_2} + |\phi_1(\lambda_-)| \|I_1^{(+)}\|_{\mathbb{L}_2} \right) \\
 &\quad + \frac{1}{2}uh_n |\phi_1(\lambda_-)| \|\tilde{I}_1\|_{\mathbb{L}_2} \\
 &\leq \left[\frac{1}{6}\gamma^2\sqrt{u} + \frac{1}{6}\gamma u\sqrt{M_1} + \frac{1}{6}u^{1.5}M_1 + \frac{15 + 10\sqrt{3} + 9\sqrt{5}}{360}\gamma^2\sigma\sqrt{h_n} \right. \\
 &\quad + \frac{15 + 5\sqrt{3} + 6\sqrt{5}}{180}\sigma u M_1 \sqrt{h_n} + \frac{3 - \sqrt{3}}{36}u^2 M_1^{1.5} h_n \\
 &\quad \left. + \frac{2 - \sqrt{3}}{36}u^{2.5} M_1^2 h_n^2 + \frac{15 - 5\sqrt{3} + 21\sqrt{5} - 11\sqrt{15}}{2160}\sigma u^2 M_1^2 h_n^{2.5} \right] \sqrt{dh_n^3} \\
 &\leq \left[\gamma^2\sqrt{u} + \gamma^2\sigma + \gamma u\sqrt{M_1} + u^{1.5}M_1 + \sigma u M_1 + u^2 M_1^{1.5} + u^{2.5}M_1^2 + \sigma u^2 M_1^2 \right] \sqrt{dh_n^3}, \\
 \|R_X^{(3)}\|_{\mathbb{L}_2} &\leq \frac{1}{24}\gamma^3 h_n^4 \|V_n\|_{\mathbb{L}_2} + \frac{1}{24}\gamma^3\sigma h_n^4 \|H_n + 6K_n\|_{\mathbb{L}_2} + \frac{1}{120}\gamma^3\sigma h_n^4 \|W_n - 12K_n\|_{\mathbb{L}_2} \\
 &\quad + \frac{1}{24}\gamma^2 u h_n^4 \|\nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{12}\gamma u h_n^4 \|\nabla^2 f(X_n) V_n\|_{\mathbb{L}_2} \\
 &\quad + \frac{1}{12}\gamma\sigma u h_n^4 \|\nabla^2 f(X_n)(H_n + 6K_n)\|_{\mathbb{L}_2} \\
 &\quad + \frac{7}{432}\gamma\sigma u h_n^4 \|\nabla^2 f(X_n)(W_n - 12K_n)\|_{\mathbb{L}_2} \\
 &\quad + \frac{1}{2}uh_n \left(|\phi_1(\lambda_+)| \|I_2^{(-)}\|_{\mathbb{L}_2} + |\phi_1(\lambda_-)| \|I_2^{(+)}\|_{\mathbb{L}_2} \right) + \frac{1}{2}uh_n |\phi_1(\lambda_-)| \|\tilde{I}_1\|_{\mathbb{L}_2} \\
 &\leq \left[\frac{1}{24}\gamma^3\sqrt{u} + \frac{1}{24}\gamma^2 u\sqrt{M_1} + \frac{1}{12}\gamma u^{1.5}M_1 + \frac{\sqrt{3}}{24}u^2 M_2 \right.
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{30 + 25\sqrt{3} + 21\sqrt{5}}{3600} \gamma^3 \sigma \sqrt{h_n} + \frac{7 + 6\sqrt{3} + 5\sqrt{5}}{432} \gamma u M_1 \sqrt{h_n} \\
 & + \frac{3 + 2\sqrt{3} + \sqrt{15}}{72} \sigma u^{1.5} M_2 \sqrt{h_n} + \frac{20 + 21\sqrt{3} + 10\sqrt{5} + 7\sqrt{15}}{1440} \sigma^2 u M_2 \\
 & + \frac{3 - \sqrt{3}}{36} u^2 M_1^{1.5} + \frac{2 - \sqrt{3}}{36} u^{2.5} M_1^2 h_n \\
 & + \left. \frac{15 - 5\sqrt{3} + 21\sqrt{5} - 11\sqrt{15}}{2160} \sigma u^2 M_1^2 h_n^{1.5} \right] dh_n^4 \\
 \leq & \left[\gamma^3 \sqrt{u} + \gamma^3 \sigma + \gamma^2 u \sqrt{M_1} + \gamma u^{1.5} M_1 + \gamma u M_1 + u^2 M_2 + \sigma u^{1.5} M_2 \right. \\
 & \left. + \sigma^2 u M_2 + u^2 M_1^{1.5} + u^{2.5} M_1^2 + \sigma u^2 M_1^2 \right] dh_n^4.
 \end{aligned}$$

Similarly for the momentum variable,

$$\begin{aligned}
 \|R_V^{(1)}\|_{\mathbb{L}_2} & \leq \frac{1}{2} \gamma^2 h_n^2 \|V_n\|_{\mathbb{L}_2} + \frac{1}{2} \gamma^2 \sigma h_n^2 \|H_n + 6K_n\|_{\mathbb{L}_2} + \frac{1}{6} \gamma^2 \sigma h_n^2 \|W_n - 12K_n\|_{\mathbb{L}_2} \\
 & \quad + \frac{1}{2} \gamma u h^2 \|\nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{2} u h_n \left(|\phi_0(\lambda_+)| \|I_1^{(-)}\|_{\mathbb{L}_2} + |\phi_0(\lambda_-)| \|I_1^{(+)}\|_{\mathbb{L}_2} \right) \\
 & \quad + \frac{1}{2} u h_n |\phi_0(\lambda_-)| \|\tilde{I}_1\|_{\mathbb{L}_2} \\
 & \leq \left[\frac{1}{2} \gamma^2 \sqrt{u} + \frac{1}{2} \gamma u \sqrt{M_1} + \frac{1}{2} u^{1.5} M_1 + \frac{2 + \sqrt{3} + \sqrt{5}}{12} \gamma^2 \sigma \sqrt{h_n} \right. \\
 & \quad + \frac{2 + \sqrt{3} + \sqrt{5}}{12} \sigma u M_1 \sqrt{h_n} + \frac{1}{6} u^2 M_1^{1.5} h_n + \frac{3 - \sqrt{3}}{36} u^{2.5} M_1^2 h_n^2 \\
 & \quad \left. + \frac{5 + 5\sqrt{5} - 2\sqrt{15}}{360} \sigma u^2 M_1^2 h_n^{2.5} \right] \sqrt{d} h_n^2 \\
 & \leq \left[\gamma^2 \sqrt{u} + \gamma^2 \sigma + \gamma u \sqrt{M_1} + u^{1.5} M_1 + \sigma u M_1 + u^2 M_1^{1.5} + u^{2.5} M_1^2 + \sigma u^2 M_1^2 \right] \sqrt{d} h_n^2, \\
 \|R_V^{(2)}\|_{\mathbb{L}_2} & \leq \frac{1}{6} \gamma^3 h_n^3 \|V_n\|_{\mathbb{L}_2} + \frac{1}{6} \gamma^3 \sigma h_n^3 \|H_n + 6K_n\|_{\mathbb{L}_2} + \frac{1}{24} \gamma^3 \sigma h_n^3 \|W_n - 12K_n\|_{\mathbb{L}_2} \\
 & \quad + \frac{1}{6} \gamma^2 u h_n^3 \|\nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{3} \gamma u h_n^3 \|\nabla^2 f(X_n) V_n\|_{\mathbb{L}_2} \\
 & \quad + \frac{1}{3} \gamma \sigma u h_n^3 \|\nabla^2 f(X_n)(H_n + 6K_n)\|_{\mathbb{L}_2} + \frac{1}{12} \gamma \sigma u h_n^3 \|\nabla^2 f(X_n)(W_n - 12K_n)\|_{\mathbb{L}_2} \\
 & \quad + \frac{1}{2} u h_n \left(|\phi_0(\lambda_+)| \|I_2^{(-)}\|_{\mathbb{L}_2} + |\phi_0(\lambda_-)| \|I_2^{(+)}\|_{\mathbb{L}_2} \right) + \frac{1}{2} u h_n |\phi_0(\lambda_-)| \|\tilde{I}_1\|_{\mathbb{L}_2} \\
 & \leq \left[\frac{1}{6} \gamma^3 \sqrt{u} + \frac{1}{6} \gamma^2 u \sqrt{M_1} + \frac{1}{3} \gamma u^{1.5} M_1 + \frac{\sqrt{3}}{3} u^2 M_2 + \frac{1}{6} u^2 M_1^{1.5} \right. \\
 & \quad + \frac{15 + 10\sqrt{3} + 9\sqrt{5}}{360} \gamma^3 \sigma \sqrt{h_n} + \frac{15 + 10\sqrt{3} + 9\sqrt{5}}{180} \gamma \sigma u M_1 \sqrt{h_n} \\
 & \quad \left. + \frac{20 + 15\sqrt{3} + 7\sqrt{15}}{60} \sigma u^{1.5} M_2 \sqrt{h_n} + \frac{3 - \sqrt{3}}{36} u^{2.5} M_1^2 h_n \right]
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{45 + 97\sqrt{3} + 21\sqrt{5} + 33\sqrt{15}}{360} \sigma^2 u M_2 h_n \\
 & + \left. \frac{5 + 5\sqrt{5} - 2\sqrt{15}}{360} \sigma u^2 M_1^2 h_n^{1.5} \right] dh_n^3 \\
 \leq & \left[\gamma^3 \sqrt{u} + \gamma^3 \sigma + \gamma^2 u \sqrt{M_1} + \gamma u^{1.5} M_1 + \gamma \sigma u M_1 + u^2 M_2 + \sigma u^{1.5} M_2 \right. \\
 & \left. + \sigma^2 u M_2 + u^2 M_1^{1.5} + u^{2.5} M_1^2 + \sigma u^2 M_1^2 \right] dh_n^3, \\
 \|R_V^{(3)}\|_{\mathbb{L}_2} \leq & \frac{1}{24} \gamma^4 h_n^4 \|V_n\|_{\mathbb{L}_2} + \frac{1}{24} \gamma^4 \sigma h_n^4 \|H_n + 6K_n\|_{\mathbb{L}_2} + \frac{1}{120} \gamma^4 \sigma h_n^4 \|W_n - 12K_n\|_{\mathbb{L}_2} \\
 & + \frac{1}{24} \gamma^3 u h_n^4 \|\nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{8} \gamma^2 u h_n^4 \|\nabla^2 f(X_n) V_n\|_{\mathbb{L}_2} \\
 & + \frac{1}{8} \gamma^2 \sigma u h_n^4 \|\nabla^2 f(X_n) (H_n + 6K_n)\|_{\mathbb{L}_2} \\
 & + \frac{7}{288} \gamma^2 \sigma u h_n^4 \|\nabla^2 f(X_n) (W_n - 12K_n)\|_{\mathbb{L}_2} \\
 & + \frac{4 - \sqrt{3}}{36} \gamma u^2 h_n^5 \|\nabla^2 f(X_n) \nabla f(X_n)\|_{\mathbb{L}_2} + \frac{1}{6} \gamma u h_n^4 \|\nabla^3 f(X_n) V_n^{\otimes 2}\|_{\mathbb{L}_2} \\
 & + \frac{1}{3} \gamma \sigma u h_n^4 \|\nabla^3 f(X_n) (V_n, H_n + 6K_n)\|_{\mathbb{L}_2} \\
 & + \frac{47}{432} \gamma \sigma u h_n^4 \|\nabla^3 f(X_n) (V_n, W_n - 12K_n)\|_{\mathbb{L}_2} \\
 & + \frac{1}{6} \sigma^2 u h_n^3 \|\nabla^3 f(X_n) (H_n + 6K_n)^{\otimes 2}\|_{\mathbb{L}_2} \\
 & + \frac{7}{288} \sigma^2 u h_n^3 \|\nabla^3 f(X_n) (W_n - 12K_n)^{\otimes 2}\|_{\mathbb{L}_2} \\
 & + \frac{1}{8} \sigma^2 u h_n^3 \|\nabla^3 f(X_n) (H_n + 6K_n, W_n - 12K_n)\|_{\mathbb{L}_2} \\
 & + \frac{1}{2} u^2 h_n^2 \left| \phi_0(\lambda_-) \phi_1\left(\frac{1}{3}\right) \right| \|\nabla^2 f(X_n) I_1^{(-)}\|_{\mathbb{L}_2} \\
 & + \frac{1}{2} u h_n \left[|\phi_0(\lambda_+)| \|I_3^{(-)}\|_{\mathbb{L}_2} + |\phi_0(\lambda_-)| \|I_3^{(+)}\|_{\mathbb{L}_2} \right] \\
 & + \frac{1}{2} u h_n |\phi_0(\lambda_-)| \|\tilde{I}_2\|_{\mathbb{L}_2} + \frac{1}{2} u h_n |\phi_0(\lambda_-)| \|\tilde{I}_3\|_{\mathbb{L}_2} \\
 \leq & \left[\frac{1}{24} \gamma^4 \sqrt{u} + \frac{1}{24} \gamma^3 u \sqrt{M_1} + \frac{1}{8} \gamma^2 u^{1.5} M_1 + \frac{\sqrt{3}}{6} \gamma u^2 M_2 \right. \\
 & + \frac{45 + 46\sqrt{3} + 21\sqrt{5} + 16\sqrt{15}}{720} \sigma^2 u M_2 + \frac{3 - \sqrt{3}}{36} u^{2.5} M_1 + \frac{1}{36} u^3 M_1 M_2 \\
 & + \frac{\sqrt{3} + 1}{12} u^{2.5} \sqrt{M_1} M_2 + \frac{\sqrt{15}}{24} u^{2.5} M_3 + \frac{30 + 25\sqrt{3} + 21\sqrt{5}}{3600} \gamma^4 \sigma \sqrt{h_n} \\
 & \left. + \frac{7 + 6\sqrt{3} + 5\sqrt{5}}{288} \gamma^2 \sigma u M_1 \sqrt{h_n} + \frac{360 + 235\sqrt{3} + 119\sqrt{5}}{2160} \gamma \sigma u^{1.5} M_2 \sqrt{h_n} \right]
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{5 + 5\sqrt{3} - 2\sqrt{15}}{360} \sigma u^2 M_1 \sqrt{h_n} + \frac{16\sqrt{3} + 9\sqrt{5} + 7\sqrt{15}}{144} \sigma u^2 M_3 \sqrt{h_n} \\
 & + \frac{30 + 15\sqrt{3} + 6\sqrt{5} + 5\sqrt{15}}{360} \sigma u^2 \sqrt{M_1} M_2 \sqrt{h_n} + \frac{4 - \sqrt{3}}{36} \gamma u^2 M_1^{1.5} h_n \\
 & + \frac{\sqrt{3} - 1}{36} u^{3.5} M_1^{1.5} M_2 h_n + \frac{\sqrt{3}}{18} u^3 M_1 M_2 h_n \\
 & + \frac{160 + 125\sqrt{3} + 70\sqrt{5} + 71\sqrt{15}}{720} \sigma^2 u^{1.5} M_3 h_n \\
 & + \frac{5\sqrt{3} - 6\sqrt{5} + 5\sqrt{15}}{1080} \sigma u^3 M_1^{1.5} M_2 h_n^{1.5} + \frac{10 + 5\sqrt{3} + 3\sqrt{15}}{360} \sigma u^{2.5} M_1 M_2 h_n^{1.5} \\
 & + \frac{2\sqrt{3} - 3}{216} u^2 M_1^2 M_2 h_n^2 + \frac{30 + 14\sqrt{3} + 9\sqrt{5} + 4\sqrt{15}}{2160} \sigma^2 u^2 M_1 M_2 h_n^2 \\
 & + \frac{-5 + 5\sqrt{3} - 11\sqrt{5} + 7\sqrt{15}}{2160} \sigma u^{1.5} M_1^2 M_2 h_n^{2.5} \\
 & + \frac{1125 + 835\sqrt{3} + 549\sqrt{5} + 371\sqrt{15}}{25920} \sigma^3 u M_3 h_n^{2.5} \\
 & + \left. \frac{-30 + 21\sqrt{3} - 6\sqrt{5} + 5\sqrt{15}}{12960} \sigma^2 u^3 M_1^2 M_2 h_n^3 \right] d^{1.5} h_n^4 \\
 \leq & \left[\gamma^4 \sqrt{u} + \gamma^4 \sigma + \gamma^3 u \sqrt{M_1} + \gamma^2 u^{1.5} M_1 + \gamma^2 \sigma u M_1 + u^{2.5} M_1 + \sigma u^2 M_1 \right. \\
 & + \gamma u^2 M_2 + \gamma \sigma u^{1.5} M_2 + \sigma^2 u M_2 + u^{2.5} M_3 + \sigma u^2 M_3 + \sigma^2 u^{1.5} M_3 + \sigma^3 u M_3 \\
 & + \gamma u^2 M_1^{1.5} + u^{2.5} \sqrt{M_1} M_2 + \sigma u^2 \sqrt{M_1} M_2 + u^3 M_1 M_2 + \sigma u^{2.5} M_1 M_2 \\
 & + \sigma^2 u^2 M_1 M_2 + u^{3.5} M_1^{1.5} M_2 + \sigma u^3 M_1^{1.5} M_2 + u^2 M_1^2 M_2 + \sigma u^{1.5} M_1^2 M_2 \\
 & \left. + \sigma^2 u^3 M_1^2 M_2 \right] d^{1.5} h_n^4.
 \end{aligned}$$

Similar to the ULD remainder bounds (Theorem 33), we can prove (72), (74) and (76) using

$$\|\mathbb{E}[R]\|_2 \leq \mathbb{E}[\|R\|_2] \leq \mathbb{E}[\|R\|_2^2]^{\frac{1}{2}} = \|R\|_{\mathbb{L}_2},$$

by Jensen's inequality. ■

We are now in a position to obtain local error bounds for QUICSORT (Theorem 10).

Theorem 36 (Local error bounds for the QUICSORT method) *Let $\{(X_n, V_n)\}_{n \geq 0}$ denote the QUICSORT method (3) and $\{(x_t, v_t)\}_{t \geq 0}$ be the underdamped Langevin diffusion (2) with the same underlying Brownian motion. Suppose the assumptions (A.1) and (A.2) hold and assume further that $(x_0, v_0) \sim \pi$, the unique stationary measure of the process. Suppose that at time t_n , we have $(X_n, V_n) = (x_{t_n}, v_{t_n})$ and let $0 < h_n \leq 1$ be fixed. Then, for any $n \geq 0$, the local strong error for the QUICSORT method at time t_{n+1} has the bound*

$$\|X_{n+1} - x_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_{s,x}^{(1)} \sqrt{d} h_n^2, \quad \|V_{n+1} - v_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_{s,v}^{(1)} \sqrt{d} h_n^2, \quad (77)$$

and the local weak error is bounded by

$$\|\mathbb{E}[X_{n+1} - x_{t_{n+1}}]\|_2 \leq C_{w,x}^{(1)} \sqrt{d} h_n^2, \quad \|\mathbb{E}[V_{n+1} - v_{t_{n+1}}]\|_2 \leq C_{w,v}^{(1)} \sqrt{d} h_n^2, \quad (78)$$

where $C_{s,x}^{(1)}, C_{s,v}^{(1)}, C_{w,x}^{(1)}, C_{w,v}^{(1)} > 0$ are constants depending on γ, u, m and M_1 .

Under the additional assumption (A.3), the local strong error at time t_{n+1} has the bound

$$\|X_{n+1} - x_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_{s,x}^{(2)} \sqrt{d} h_n^3, \quad \|V_{n+1} - v_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_{s,v}^{(2)} d h_n^3, \quad (79)$$

and the local weak error is bounded by

$$\|\mathbb{E}[X_{n+1} - x_{t_{n+1}}]\|_2 \leq C_{w,x}^{(2)} \sqrt{d} h_n^3, \quad \|\mathbb{E}[V_{n+1} - v_{t_{n+1}}]\|_2 \leq C_{w,v}^{(2)} d h_n^3, \quad (80)$$

where $C_{s,x}^{(2)}, C_{s,v}^{(2)}, C_{w,x}^{(2)}, C_{w,v}^{(2)} > 0$ are constants depending on γ, u, m, M_1 and M_2 . Finally, under the additional assumption (A.4), the local strong error at time t_{n+1} has the bound

$$\|X_{n+1} - x_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_{s,x}^{(3)} d h_n^{3.5}, \quad \|V_{n+1} - v_{t_{n+1}}\|_{\mathbb{L}_2} \leq C_{s,v}^{(3)} d^{1.5} h_n^{3.5}, \quad (81)$$

and the local weak error is bounded by

$$\|\mathbb{E}[X_{n+1} - x_{t_{n+1}}]\|_2 \leq C_{w,x}^{(3)} d h_n^4, \quad \|\mathbb{E}[V_{n+1} - v_{t_{n+1}}]\|_2 \leq C_{w,v}^{(3)} d^{1.5} h_n^4, \quad (82)$$

where $C_{s,x}^{(3)}, C_{s,v}^{(3)}, C_{w,x}^{(3)}, C_{w,v}^{(3)} > 0$ are constants depending on γ, u, m, M_1, M_2 and M_3 .

Proof We first note that the Taylor expansions of ULD and QUICSORT are identical for all terms that are $O(h^3)$ or larger. Then, under the assumptions (A.1) and (A.2), we have

$$\begin{aligned} \|\mathbb{E}[X_{n+1} - x_{t_{n+1}}]\|_2 &= \|X_{n+1} - x_{t_{n+1}}\|_{\mathbb{L}_2} \leq \|R_X^{(1)}\|_{\mathbb{L}_2} + \|R_x^{(1)}\|_{\mathbb{L}_2} \\ &\leq [C_X^{(1)} + C_x^{(1)}] \sqrt{d} h_n^2, \\ \|\mathbb{E}[V_{n+1} - v_{t_{n+1}}]\|_2 &= \|V_{n+1} - v_{t_{n+1}}\|_{\mathbb{L}_2} \leq \|R_V^{(1)}\|_{\mathbb{L}_2} + \|R_v^{(1)}\|_{\mathbb{L}_2} \\ &\leq [C_V^{(1)} + C_v^{(1)}] \sqrt{d} h_n^2, \end{aligned}$$

and, if we further assume (A.3), then

$$\begin{aligned} \|\mathbb{E}[X_{n+1} - x_{t_{n+1}}]\|_2 &= \|X_{n+1} - x_{t_{n+1}}\|_{\mathbb{L}_2} \leq \|R_X^{(2)}\|_{\mathbb{L}_2} + \|R_x^{(2)}\|_{\mathbb{L}_2} \\ &\leq [C_X^{(2)} + C_x^{(2)}] \sqrt{d} h_n^3, \\ \|\mathbb{E}[V_{n+1} - v_{t_{n+1}}]\|_2 &= \|V_{n+1} - v_{t_{n+1}}\|_{\mathbb{L}_2} \leq \|R_V^{(2)}\|_{\mathbb{L}_2} + \|R_v^{(2)}\|_{\mathbb{L}_2} \\ &\leq [C_V^{(2)} + C_v^{(2)}] d h_n^3. \end{aligned}$$

In the case where we additionally assume (A.4), we note that the triple time integral of the Brownian motion is centred, and we can write it in terms of W_n , H_n , K_n , M_n using (33). Moreover, the other stochastic integrals are centred and we can bound the following terms,

$$\begin{aligned}
 & \left\| \sigma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^3 f(x_{t_n})(v_{r_2}, W_{t_n, r_3}) dr_3 dr_2 dr_1 \right\|_{\mathbb{L}_2} \\
 & \leq \sigma u M_2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \|v_{r_2}\|_{\mathbb{L}_4} \|W_{t_n, r_3}\|_{\mathbb{L}_4} dr_3 dr_2 dr_1 \\
 & \leq \sqrt{3} \sigma u^{1.5} M_2 d \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} (r_3 - t_n)^{\frac{1}{2}} dr_3 dr_2 dr_1 \\
 & \leq \frac{8\sqrt{3}}{105} \sigma u^{1.5} M_2 d h_n^{3.5},
 \end{aligned}$$

and

$$\begin{aligned}
 & \left\| \sigma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \nabla^3 f(x_{t_n})(v_{t_n}, (r_2 - t_n)W_{t_n, r_2}) dr_2 dr_1 \right\|_{\mathbb{L}_2} \\
 & \leq \sigma u M_2 \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} (r_2 - t_n) \|v_{t_n}\|_{\mathbb{L}_4} \|W_{t_n, r_2}\|_{\mathbb{L}_4} dr_2 dr_1 \\
 & \leq \sqrt{3} \sigma u^{1.5} M_2 d \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} (r_2 - t_n)^{\frac{3}{2}} dr_2 dr_1 \\
 & \leq \frac{4\sqrt{3}}{35} \sigma u^{1.5} M_2 d h_n^{3.5},
 \end{aligned}$$

by Fubini's theorem along with the bounds (31), (34) and (40). We can show (81) and (82) for the position variable as follows

$$\begin{aligned}
 \|X_{n+1} - x_{t_{n+1}}\|_{\mathbb{L}_2} & \leq \left\| \sigma(\gamma^2 - u\nabla^2 f(x_{t_n})) \left(\frac{1}{60} h_n^3 H_n - h_n^3 M_n \right) \right\|_{\mathbb{L}_2} + \|R_X^{(3)}\|_{\mathbb{L}_2} + \|R_x^{(3)}\|_{\mathbb{L}_2} \\
 & \leq \left[\frac{7\sqrt{3} + 3\sqrt{7}}{2520} \sigma(\gamma^2 - um) + C_X^{(3)} \sqrt{h_n} + C_x^{(3)} \sqrt{h_n} \right] d h_n^{3.5} \\
 & \leq \left[\sigma(\gamma^2 - um) + C_X^{(3)} + C_x^{(3)} \right] d h_n^{3.5},
 \end{aligned}$$

$$\begin{aligned}
 \|\mathbb{E}[X_{n+1} - x_{t_{n+1}}]\|_2 & \leq \|R_X^{(3)}\|_{\mathbb{L}_2} + \|R_x^{(3)}\|_{\mathbb{L}_2} \\
 & \leq \left[C_X^{(3)} + C_x^{(3)} \right] d h_n^4,
 \end{aligned}$$

and for the momentum variable,

$$\begin{aligned}
 \|V_{n+1} - v_{t_{n+1}}\|_{\mathbb{L}_2} & \leq \left\| \sigma(\gamma^2 - 2u\nabla^2 f(x_{t_n})) \left(\frac{1}{60} h_n^3 H_n - h_n^3 M_n \right) \right\|_{\mathbb{L}_2} \\
 & \quad + \left\| \sigma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \int_{t_n}^{r_2} \nabla^3 f(x_{t_n})(v_{r_2}, W_{t_n, r_3}) dr_3 dr_2 dr_1 \right\|_{\mathbb{L}_2}
 \end{aligned}$$

$$\begin{aligned}
 & + \left\| \sigma u \int_{t_n}^{t_{n+1}} \int_{t_n}^{r_1} \nabla^3 f(x_{t_n})(v_{t_n}, (r_2 - t_n)W_{t_n, r_2}) dr_2 dr_1 \right\|_{\mathbb{L}_2} \\
 & + \|R_V^{(3)}\|_{\mathbb{L}_2} + \|R_v^{(3)}\|_{\mathbb{L}_2} \\
 & \leq \left[\frac{7\sqrt{3} + 3\sqrt{7}}{2520} \sigma(\gamma^2 - um) + \frac{12\sqrt{3}}{35} \sigma u^{1.5} M_2 \right. \\
 & \quad \left. + C_V^{(3)} \sqrt{h_n} + C_v^{(3)} \sqrt{h_n} \right] d^{1.5} h_n^{3.5} \\
 & \leq \left[\sigma(\gamma^2 - um) + \sigma u^{1.5} M_2 + C_V^{(3)} + C_v^{(3)} \right] d^{1.5} h_n^{3.5},
 \end{aligned}$$

$$\|\mathbb{E}[V_{n+1} - v_{t_{n+1}}]\|_2 \leq \|R_V^{(3)}\|_{\mathbb{L}_2} + \|R_v^{(3)}\|_{\mathbb{L}_2} \leq \left[C_V^{(3)} + C_v^{(3)} \right] d^{1.5} h_n^4,$$

where we have used Minkowski's inequality and the bounds on the remainder terms. \blacksquare

Corollary 37 *Under the same settings as used in Theorem 36, assuming (A.1) and (A.2), we have the additional bounds*

$$\|\mathbb{E}[X_{n+1} - x_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \leq C_{w,x}^{(1)} \sqrt{d} h_n^2, \quad \|\mathbb{E}[V_{n+1} - v_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \leq C_{w,v}^{(1)} \sqrt{d} h_n^2, \quad (83)$$

where $C_{w,x}^{(1)}, C_{w,v}^{(1)} > 0$ are constants depending on γ, u, m and M_1 .

Under the additional assumption (A.3), we have the bounds

$$\|\mathbb{E}[X_{n+1} - x_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \leq C_{w,x}^{(2)} d h_n^3, \quad \|\mathbb{E}[V_{n+1} - v_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \leq C_{w,v}^{(2)} d h_n^3, \quad (84)$$

where $C_{w,x}^{(2)}, C_{w,v}^{(2)} > 0$ are constants depending on γ, u, m, M_1 and M_2 .

Finally, under the additional assumption (A.4), we have the bounds

$$\|\mathbb{E}[X_{n+1} - x_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \leq C_{w,x}^{(3)} d^{1.5} h_n^4, \quad \|\mathbb{E}[V_{n+1} - v_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \leq C_{w,v}^{(3)} d^{1.5} h_n^4, \quad (85)$$

where $C_{w,x}^{(3)}, C_{w,v}^{(3)} > 0$ are constants depending on γ, u, m, M_1, M_2 and M_3 .

Proof Notice that the Taylor expansion for the ULD and the QUICSORT method are identical up to order h^3 , and the only mismatch are the stochastic integrals of order $h^{3.5}$. However, these integrals have a conditional mean of zero, and thus do not contribute to the above errors. Moreover, for both the position and momentum variables' remainder terms, under any of the smoothness assumptions,

$$\|\mathbb{E}[R | \mathcal{F}_n]\|_{\mathbb{L}_2} = \left(\mathbb{E} \left[\|\mathbb{E}[R | \mathcal{F}_n]\|_2^2 \right] \right)^{\frac{1}{2}} \leq \left(\mathbb{E} \left[\mathbb{E}[\|R\|_2^2 | \mathcal{F}_n] \right] \right)^{\frac{1}{2}} = \|R\|_{\mathbb{L}_2},$$

by Jensen's inequality and the tower property of conditional expectations. The result now follows by applying the same logic as used in the proof of Theorem 36 to establish the local weak error bounds. \blacksquare

Remark 38 *The bounds for the \mathbb{L}_2 norm of the absolute conditional mean are identical to the bounds for the local weak error. This will allow us to use them interchangeably when proving the global error result (Theorem 7).*

Appendix D. Global error bounds

In this section, we prove the global error bounds for the QUICSORT method (Theorem 7). To do so, we use the methodology outlined in Section 3. We will also use the following theorem.

Theorem 39 *Consider two approximations $Q = (X, V)$ and $\widehat{Q} = (Y, U)$ under the mapping $Q_n \rightarrow Q_{n+1}$ and $\widehat{Q}_n \rightarrow \widehat{Q}_{n+1}$ defined in equation (3), driven by the same Brownian motion but evolved from the initial conditions $Q_n = (X_n, V_n)$ and $\widehat{Q}_n = (Y_n, U_n)$ respectively. Suppose assumption (A.2) holds and let $0 < h_n \leq 1$. Then for all $n \geq 0$*

$$\|(Q_{n+1} - Q_n) - (\widehat{Q}_{n+1} - \widehat{Q}_n)\|_{\mathbb{L}_2} \leq C_q \|Q_n - \widehat{Q}_n\|_{\mathbb{L}_2} h_n, \quad (86)$$

where $C_q > 0$ is a constant depending on γ , u , and M_1 .

Proof We first rewrite the QUICSORT method (3) as

$$\begin{aligned} X_{n+1} &= X_n + \phi_1(1)(V_n + \sigma(H_n + 6K_n)) + \phi_2(1)C_n \\ &\quad - \frac{1}{2}uh_n \left(\phi_1(\lambda_+) \nabla f(X_n^{(1)}) + \phi_1(\lambda_-) \nabla f(X_n^{(2)}) \right), \\ V_{n+1} &= \phi_0(1) \left(V_n + \sigma(H_n + 6K_n) \right) + \sigma(H_n - 6K_n) + \phi_1(1)C_n \\ &\quad - \frac{1}{2}uh_n \left(\phi_0(\lambda_+) \nabla f(X_n^{(1)}) + \phi_0(\lambda_-) \nabla f(X_n^{(2)}) \right), \end{aligned}$$

where

$$\begin{aligned} X_n^{(1)} &= X_n + \phi_1(\lambda_-)(V_n + \sigma(H_n + 6K_n)) + \phi_2(\lambda_-)C_n, \\ X_n^{(2)} &= X_n + \phi_1(\lambda_+)(V_n + \sigma(H_n + 6K_n)) + \phi_2(\lambda_+)C_n - h_n \phi_1 \left(\frac{1}{3} \right) u \nabla f(X_n^{(1)}). \end{aligned}$$

Since the approximations Q and \widehat{Q} use the same Brownian motion, we can show that

$$\begin{aligned} &\|(X_{n+1} - X_n) - (Y_{n+1} - Y_n)\|_{\mathbb{L}_2} \\ &\leq |\phi_1(1)| \|V_n - U_n\|_{\mathbb{L}_2} + \frac{1}{2}uh_n |\phi_1(\lambda_+)| \|\nabla f(X_n^{(1)}) - \nabla f(Y_n^{(1)})\|_{\mathbb{L}_2} \\ &\quad + \frac{1}{2}uh_n |\phi_1(\lambda_-)| \|\nabla f(X_n^{(2)}) - \nabla f(Y_n^{(2)})\|_{\mathbb{L}_2} \\ &\leq \frac{1}{2}uM_1 h_n \left(|\phi_1(\lambda_+)| + |\phi_1(\lambda_-)| + uM_1 h_n \left| \phi_1 \left(\frac{1}{3} \right) \phi_1(\lambda_-) \right| \right) \|X_n - Y_n\|_{\mathbb{L}_2} \\ &\quad + \left(|\phi_1(1)| + uM_1 h_n |\phi_1(\lambda_+) \phi_1(\lambda_-)| + \frac{1}{2}u^2 M_1^2 h_n^2 \left| \phi_1 \left(\frac{1}{3} \right) \phi_1^2(\lambda_-) \right| \right) \|V_n - U_n\|_{\mathbb{L}_2} \\ &\leq (1 + uM_1 + u^2 M_1^2) \|Q_n - \widehat{Q}_n\|_{\mathbb{L}_2} h_n, \end{aligned}$$

and, similarly for the momentum variable, we have

$$\begin{aligned}
 & \| (V_{n+1} - V_n) - (U_{n+1} - U_n) \|_{\mathbb{L}_2} \\
 & \leq |\phi_0(1) - 1| \| V_n - U_n \|_{\mathbb{L}_2} + \frac{1}{2} u h_n |\phi_0(\lambda_+)| \| \nabla f(X_n^{(1)}) - \nabla f(Y_n^{(1)}) \|_{\mathbb{L}_2} \\
 & \quad + \frac{1}{2} u h_n |\phi_0(\lambda_-)| \| \nabla f(X_n^{(2)}) - \nabla f(Y_n^{(2)}) \|_{\mathbb{L}_2} \\
 & \leq \frac{1}{2} u M_1 h_n \left[|\phi_0(\lambda_+)| + |\phi_0(\lambda_-)| + u M_1 h_n \left| \phi_0(\lambda_-) \phi_1\left(\frac{1}{3}\right) \right| \right] \| X_n - Y_n \|_{\mathbb{L}_2} \\
 & \quad + \left(|\phi_0(1) - 1| + \frac{1}{2} u M_1 h_n (|\phi_0(\lambda_+) \phi_1(\lambda_-)| + |\phi_0(\lambda_-) \phi_1(\lambda_+)|) \right. \\
 & \quad \left. + \frac{1}{2} u^2 M_1^2 h_n^2 \left| \phi_0(\lambda_-) \phi_1\left(\frac{1}{3}\right) \phi_1(\lambda_-) \right| \right) \| V_n - U_n \|_{\mathbb{L}_2} \\
 & \leq (\gamma + 2uM_1 + u^2M_1^2) \| Q_n - \widehat{Q}_n \|_{\mathbb{L}_2} h_n.
 \end{aligned}$$

The result follows by summing the bounds for the position and momentum variables. \blacksquare

Finally, we will show the main result of our paper (global error bounds for QUICSORT).

Theorem 40 (Global error bounds for the QUICSORT method) *Let $\{(X_n, V_n)\}_{n \geq 0}$ denote the QUICSORT method (3) and let $\{(x_t, v_t)\}_{t \geq 0}$ be the underdamped Langevin diffusion (2) driven by the same underlying Brownian motion. Suppose assumptions (A.1) and (A.2) hold. Further suppose that $(x_0, v_0) \sim \pi$, the unique stationary measure of the diffusion process and that the approximation and the SDE solution have the same initial velocity $V_0 = v_0 \sim \mathcal{N}(0, uI_d)$. Let $0 < h \leq 1$ be fixed and $\gamma \geq 2\sqrt{uM_1}$. Then for all $n \geq 0$, the global error at time t_n has the bound*

$$\| X_n - x_{t_n} \|_{\mathbb{L}_2} \leq c e^{-n\beta h} \| X_0 - x_0 \|_{\mathbb{L}_2} + C_1 \sqrt{d} h,$$

where $c, \beta, C_1 > 0$ are constants depending on γ, u, m and M_1 .

Under the additional assumption (A.3), the global error at time t_n satisfies the bound

$$\| X_n - x_{t_n} \|_{\mathbb{L}_2} \leq c e^{-n\beta h} \| X_0 - x_0 \|_{\mathbb{L}_2} + C_2 d h^2,$$

where $C_2 > 0$ is a constant depending on γ, u, m, M_1 and M_2 .

Finally, under the additional assumption (A.4), the global error at time t_n has the bound

$$\| X_n - x_{t_n} \|_{\mathbb{L}_2} \leq c e^{-\frac{1}{2}n\beta h} \| X_0 - x_0 \|_{\mathbb{L}_2} + C_3 d^{1.5} h^3,$$

where $C_3 > 0$ is a constant depending on γ, u, m, M_1, M_2 and M_3 .

Proof The approach for this proof will follow the strategy that was outlined in Section 3. We will write $Q_n = (X_n, V_n)^\top$, $q_{t_n} = (x_{t_n}, v_{t_n})^\top$ and define the coordinate transformation

$$A := \begin{pmatrix} wI_d & I_d \\ zI_d & I_d \end{pmatrix},$$

where $w \in [0, \frac{1}{2}\gamma)$, $z = \gamma - w$. We will define the errors with and without this transformation,

$$\begin{aligned} e_n &:= \|Q_n - q_{t_n}\|_{\mathbb{L}_2}, \\ \tilde{e}_n &:= \|A(Q_n - q_{t_n})\|_{\mathbb{L}_2}. \end{aligned}$$

Let $\sigma_{\max} = \|A\|_2$ and σ_{\min} denote the smallest and largest singular values of A respectively. Since A is invertible, the singular value σ_{\min} is non-zero and will satisfy $\sigma_{\min} = \|A^{-1}\|_2^{-1}$. Let $\kappa = \frac{\sigma_{\max}}{\sigma_{\min}}$ denote the condition number of A . We can use \tilde{e}_n to bound the error e_n as

$$\sigma_{\min} e_n \leq \tilde{e}_n \leq \sigma_{\max} e_n.$$

We now consider a third process $\widehat{Q} = (\widehat{X}, \widehat{V})^\top$ where $(\widehat{X}_{n+1}, \widehat{V}_{n+1})$ is obtained by applying the QUICSORT method (3) to the ULD diffusion at time t_n with the same underlying Brownian motion as both Q and q . This allows us to estimate the local error between the coordinate-changed approximation AQ_{n+1} and diffusion process $Aq_{t_{n+1}}$ at time t_{n+1} in terms of the same error at time t_n . By Theorem 10, we can bound the local strong error as

$$\|\widehat{Q}_n - q_{t_n}\|_{\mathbb{L}_2} \leq C_s d^q h^{p_s},$$

and the local weak error as

$$\|\mathbb{E}[\widehat{Q}_n - q_{t_n}]\|_2 \leq C_w d^q h^{p_w},$$

where $C_s, C_w, q, p_s, p_w > 0$ are the following constants (which do not depend on d and h).

Assumptions	C_s	C_w	q	p_s	p_w
(A.1), (A.2)	$C_{s,x}^{(1)} + C_{s,v}^{(1)}$	$C_{w,x}^{(1)} + C_{w,v}^{(1)}$	0.5	2	2
(A.1), (A.2), (A.3)	$C_{s,x}^{(2)} + C_{s,v}^{(2)}$	$C_{w,x}^{(2)} + C_{w,v}^{(2)}$	1	3	3
(A.1), (A.2), (A.3), (A.4)	$C_{s,x}^{(3)} + C_{s,v}^{(3)}$	$C_{w,x}^{(3)} + C_{w,v}^{(3)}$	1.5	3.5	4

In the above table, we also note that $p_s + \frac{1}{2} \geq p_w$. We now define the following variables

$$\begin{aligned} y_n &:= A(Q_n - q_{t_n}) = \begin{pmatrix} (wX_n + V_n) - (wx_{t_n} + vt_n) \\ (zX_n + V_n) - (zx_{t_n} + vt_n) \end{pmatrix}, \\ \tilde{y}_n &:= A(Q_n - \widehat{Q}_n) = \begin{pmatrix} (wX_n + V_n) - (w\widehat{X}_n + \widehat{V}_n) \\ (zX_n + V_n) - (z\widehat{X}_n + \widehat{V}_n) \end{pmatrix}, \\ \widehat{y}_n &:= A(\widehat{Q}_n - q_{t_n}) = \begin{pmatrix} (w\widehat{X}_n + \widehat{V}_n) - (wx_{t_n} + vt_n) \\ (z\widehat{X}_n + \widehat{V}_n) - (zx_{t_n} + vt_n) \end{pmatrix}, \end{aligned}$$

and note that $y_n = \tilde{y}_n + \widehat{y}_n$. Thus, the coordinate-transformed error \tilde{e}_n can be estimated as

$$\tilde{e}_{n+1} = \|y_{n+1}\|_{\mathbb{L}_2} = \|\tilde{y}_{n+1} + \widehat{y}_{n+1}\|_{\mathbb{L}_2} \leq \|\tilde{y}_{n+1}\|_{\mathbb{L}_2} + \|\widehat{y}_{n+1}\|_{\mathbb{L}_2}.$$

We bound the first term using the contractivity of the QUICSORT method (Theorem 29),

$$\|\tilde{y}_{n+1}\|_{\mathbb{L}_2} \leq \sqrt{1 - 2\alpha h} \tilde{e}_n,$$

and the second term can be bounded using the local strong error,

$$\|\hat{y}_{n+1}\|_{\mathbb{L}_2} \leq \sigma_{\max} C_s d^q h^{p_s}.$$

Therefore, we obtain the following estimate of the coordinate-transformed error,

$$\tilde{e}_{n+1} \leq \sqrt{1 - 2\alpha h} \tilde{e}_n + \sigma_{\max} C_s d^q h^{p_s}. \quad (87)$$

By setting $w = 0$ and using Remark 30, we have $\alpha > 0$ provided h is sufficiently small. Similarly, by assuming h is sufficiently small so that $\alpha h \leq \frac{1}{2}$, we obtain the below inequality.

$$\sqrt{1 - 2\alpha h} \leq \sqrt{1 - 2\alpha h + \alpha^2 h^2} = 1 - \alpha h.$$

Therefore, using this inequality and by applying the inequality (87) n times to \tilde{e}_n , we obtain

$$\begin{aligned} \tilde{e}_n &\leq (\sqrt{1 - 2\alpha h})^n \tilde{e}_0 + \sum_{k=0}^{n-1} (\sqrt{1 - 2\alpha h})^k \sigma_{\max} C_s d^q h^{p_s} \\ &\leq (1 - \alpha h)^n \tilde{e}_0 + \sum_{k=0}^{n-1} (1 - \alpha h)^k \sigma_{\max} C_s d^q h^{p_s} \\ &= (1 - \alpha h)^n \tilde{e}_0 + \left(\frac{1 - (1 - \alpha h)^n}{1 - (1 - \alpha h)} \right) \sigma_{\max} C_s d^q h^{p_s} \\ &\leq (1 - \alpha h)^n \tilde{e}_0 + \left(\frac{1}{1 - (1 - \alpha h)} \right) \sigma_{\max} C_s d^q h^{p_s} \\ &\leq e^{-n\alpha h} \tilde{e}_0 + \frac{1}{\alpha} \sigma_{\max} C_s d^q h^{p_s-1}. \end{aligned}$$

Since $e_n \leq \frac{1}{\sigma_{\min}} \tilde{e}_n$ and $\tilde{e}_0 \leq \sigma_{\max} e_0$, it follows that the error e_n can be bounded as

$$e_n \leq \kappa_A e^{-n\alpha h} e_0 + \kappa_A \left(\frac{1}{\alpha} C_s \right) d^q h^{p_s-1}, \quad (88)$$

for $n \geq 0$. Using $e_0 = \|X_0 - x_0\|_{\mathbb{L}_2}$ and $\|X_n - x_{t_n}\|_{\mathbb{L}_2} \leq e_n$, the inequality (88) gives the first two bounds (under the assumptions ((A.1), (A.2)) and ((A.1), (A.2), (A.3)) respectively). However, the estimate would only be $O(h^{2.5})$ in the third case. Thus, we now expand \tilde{e}_n as

$$\begin{aligned} \tilde{e}_{n+1}^2 &= \|\tilde{y}_{n+1} + \hat{y}_{n+1}\|_{\mathbb{L}_2}^2 \\ &= \|\tilde{y}_{n+1}\|_{\mathbb{L}_2}^2 + \|\hat{y}_{n+1}\|_{\mathbb{L}_2}^2 + 2\mathbb{E}[\langle \tilde{y}_{n+1}, \hat{y}_{n+1} \rangle]. \end{aligned}$$

We bound the first term using the contractivity of the QUICSORT method (Theorem 29),

$$\|\tilde{y}_{n+1}\|_{\mathbb{L}_2}^2 \leq (1 - 2\alpha h) \tilde{e}_n^2,$$

and the second term can be bounded using the local strong error,

$$\|\widehat{y}_{n+1}\|_{\mathbb{L}_2}^2 \leq \sigma_{\max}^2 C_s^2 d^{2q} h^{2p_s}.$$

Let \mathcal{F}_n be the filtration underlying the processes at time t_n . Then, by the tower property of conditional expectation and the Cauchy-Schwarz inequality, we can bound the third term

$$\begin{aligned} \mathbb{E}[\langle \widetilde{y}_{n+1}, \widehat{y}_{n+1} \rangle] &= \mathbb{E}[\langle y_n, \widehat{y}_{n+1} \rangle] + \mathbb{E}[\langle \widetilde{y}_{n+1} - y_n, \widehat{y}_{n+1} \rangle] \\ &= \mathbb{E}[\langle y_n, \mathbb{E}[\widehat{y}_{n+1} | \mathcal{F}_n] \rangle] + \mathbb{E}[\langle \widetilde{y}_{n+1} - y_n, \widehat{y}_{n+1} \rangle] \\ &\leq \|y_n\|_{\mathbb{L}_2} \|\mathbb{E}[\widehat{y}_{n+1} | \mathcal{F}_n]\|_{\mathbb{L}_2} + \|\widetilde{y}_{n+1} - y_n\|_{\mathbb{L}_2} \|\widehat{y}_{n+1}\|_{\mathbb{L}_2} \\ &\leq \widetilde{e}_n \|A\|_2 \|\mathbb{E}[\widehat{Q}_{n+1} - q_{t_{n+1}} | \mathcal{F}_n]\|_{\mathbb{L}_2} \\ &\quad + \|A\|_2^2 \|(Q_{n+1} - \widehat{Q}_{n+1}) - (Q_n - q_{t_n})\|_{\mathbb{L}_2} \|\widehat{Q}_{n+1} - q_{t_{n+1}}\|_{\mathbb{L}_2}. \end{aligned}$$

Using the local strong error bounds for QUICSORT, Theorem 39 and Corollary 37, we have

$$\begin{aligned} \mathbb{E}[\langle \widetilde{y}_{n+1}, \widehat{y}_{n+1} \rangle] &\leq \sigma_{\max} \widetilde{e}_n C_w d^q h^{p_w} + \sigma_{\max}^2 C_q e_n h \cdot C_s d^q h^{p_s} \\ &\leq \sigma_{\max} \widetilde{e}_n (C_w + \kappa_A C_q C_s) d^q h^{p_w}, \end{aligned}$$

where we also used the assumption that $h \leq 1$. Putting everything together and applying the inequality (43) in Theorem 25 with $c = \frac{1}{\alpha h}$, we obtain

$$\begin{aligned} \widetilde{e}_{n+1}^2 &\leq (1 - 2\alpha h) \widetilde{e}_n^2 + \sigma_{\max}^2 C_s^2 d^{2q} h^{2p_s} + 2\sigma_{\max} \widetilde{e}_n (C_w + \kappa_A C_q C_s) d^q h^{p_w} \\ &\leq (1 - 2\alpha h) \widetilde{e}_n^2 + \sigma_{\max}^2 C_s^2 d^{2q} h^{2p_s} + \alpha h \widetilde{e}_n^2 + \frac{1}{\alpha} \sigma_{\max}^2 (C_w + \kappa_A C_q C_s)^2 d^{2q} h^{2p_w-1} \\ &\leq (1 - \alpha h) \widetilde{e}_n^2 + \sigma_{\max}^2 \left(C_s^2 + \frac{1}{\alpha} (C_w + \kappa_A C_q C_s)^2 \right) d^{2q} h^{2p_w-1}. \end{aligned}$$

Applying the above inequality n times to \widetilde{e}_n gives

$$\begin{aligned} \widetilde{e}_n^2 &\leq (1 - \alpha h)^n \widetilde{e}_0^2 + \sigma_{\max}^2 \left(C_s^2 + \frac{1}{\alpha} (C_w + \kappa_A C_q C_s)^2 \right) d^{2q} h^{2p_w-1} \sum_{i=0}^{n-1} (1 - \alpha h)^i \\ &= (1 - \alpha h)^n \widetilde{e}_0^2 + \sigma_{\max}^2 \left(C_s^2 + \frac{1}{\alpha} (C_w + \kappa_A C_q C_s)^2 \right) d^{2q} h^{2p_w-1} \left(\frac{1 - (1 - \alpha h)^n}{\alpha h} \right) \\ &\leq e^{-n\alpha h} \widetilde{e}_0^2 + \sigma_{\max}^2 \left(\frac{1}{\alpha} C_s^2 + \frac{1}{\alpha^2} (C_w + \kappa_A C_q C_s)^2 \right) d^{2q} h^{2p_w-2}. \end{aligned}$$

Like earlier, we set $w = 0$ and used Remark 30 so that $0 < \alpha h \leq 1$ if h is sufficiently small. Taking a square root on both sides of the inequality and using $\sqrt{a^2 + b^2} \leq a + b$ for $a, b \geq 0$,

$$\widetilde{e}_n \leq e^{-\frac{1}{2}n\alpha h} \widetilde{e}_0 + \sigma_{\max} \left(\frac{1}{\sqrt{\alpha}} C_s + \frac{1}{\alpha} (C_w + \kappa_A C_q C_s) \right) d^q h^{p_w-1}.$$

Using $e_n \leq \frac{1}{\sigma_{\min}} \widetilde{e}_n$ and $\widetilde{e}_0 \leq \sigma_{\max} e_0$, we obtain

$$e_n \leq \kappa_A e^{-\frac{1}{2}n\alpha h} e_0 + \kappa_A \left(\frac{1}{\sqrt{\alpha}} C_s + \frac{1}{\alpha} (C_w + \kappa_A C_q C_s) \right) d^q h^{p_w-1},$$

for all $n \geq 0$. The third result now follows since $e_0 = \|X_0 - x_0\|_{\mathbb{L}_2}$ and $\|X_n - x_{t_n}\|_{\mathbb{L}_2} \leq e_n$. \blacksquare