A uniformly accurate scheme for the numerical integration of penalized Langevin dynamics

Adrien Laurent[1]

September 1, 2022

Abstract

In molecular dynamics, penalized overdamped Langevin dynamics are used to model the motion of a set of particles that follow constraints up to a parameter ε. The most used schemes for simulating these dynamics are the Euler integrator in $\mathbb{R}^d$ and the constrained Euler integrator. Both have weak order one of accuracy, but work properly only in specific regimes depending on the size of the parameter ε. We propose in this paper a new consistent method with an accuracy independent of ε for solving penalized dynamics on a manifold of any dimension. Moreover, this method converges to the constrained Euler scheme when ε goes to zero. The numerical experiments confirm the theoretical findings, in the context of weak convergence and for the invariant measure, on a torus and on the orthogonal group in high dimension and high codimension.

Keywords: constrained stochastic differential equations, penalized Langevin dynamics, manifolds, uniform accuracy, invariant measure.

AMS subject classification (2020): 60H35, 70H45, 37M25.

1 Introduction

In molecular dynamics, the overdamped Langevin equation in $\mathbb{R}^d$ is often used for modeling the behavior of a large set of particles in a high friction regime. It is given by

$$dX(t) = f(X(t))dt + \sigma dW(t), \quad X(0) = X_0, \tag{1.1}$$

where f is a smooth Lipschitz function (typically of the form $f = -\nabla V$ for V a smooth potential), $\sigma > 0$ is a constant scalar, and W is a standard d-dimensional Brownian motion in $\mathbb{R}^d$ on a probability space equipped with a filtration $(\mathcal{F}_t)$ and fulfilling the usual assumptions. If the particles are subject to smooth constraints $\zeta \colon \mathbb{R}^d \to \mathbb{R}^q$, such as strong covalent bonds between atoms or fixed angles in molecules, the dynamics follow the constrained overdamped Langevin equation

$$dX^0(t) = \Pi_{\mathcal{M}}(X^0(t))f(X^0(t))dt + \sigma\Pi_{\mathcal{M}}(X^0(t)) \circ dW(t), \quad X^0(0) = X_0 \in \mathcal{M}, \tag{1.2}$$

where the solution lies on the manifold $\mathcal{M} = \{x \in \mathbb{R}^d, \zeta(x) = 0\}$ with codimension q thanks to $\Pi_{\mathcal{M}} \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$, the orthogonal projection on the tangent bundle of the manifold $\mathcal{M}$.

In physical applications, constrained systems are often used as a limit model for stiff equations. For instance, in the dynamics of a diatomic molecule, the distance between the

---

two atoms oscillates around an average length, called the bond length (see, for instance, [34, Sect. 1.2.1] on the interactions of particles). One can work with a simpler constrained dynamics where the distance between the atoms is fixed as a constraint, or with the original (possibly stiff) dynamics in $\mathbb{R}^d$. We refer the reader to [41, 34], and references therein, for discussions on the use of constraints and penalizations in molecular dynamics. For overdamped Langevin dynamics (1.1), choosing a function $\widetilde{f} = -\nabla\widetilde{V}$ with the potential

$$\widetilde{V} = V - \frac{\sigma^2}{4}\ln(\det(G)) + \frac{1}{2\varepsilon}|\zeta|^2$$

gives penalized Langevin dynamics [16] of the form

$$dY^\varepsilon(t) = f(Y^\varepsilon(t))dt + \sigma dW(t) + \frac{\sigma^2}{4}\nabla\ln(\det(G))(Y^\varepsilon(t))dt - \frac{1}{\varepsilon}(g\zeta)(Y^\varepsilon(t))dt, \qquad (1.3)$$

where we fix $Y^\varepsilon(0) = X_0$, the parameter $\varepsilon > 0$ is fixed with arbitrary size, $f = -\nabla V$, $g = \nabla\zeta\colon \mathbb{R}^d \to \mathbb{R}^{d\times q}$, and $G = g^T g\colon \mathbb{R}^d \to \mathbb{R}^{q\times q}$ is the Gram matrix. It was shown in [16, Appx. C] that the solution Y^ε of (1.3) converges strongly to the solution X^0 of the constrained dynamics (1.2) if $X_0 \in \mathcal{M}$. The additional term $\frac{\sigma^2}{4}\nabla\ln(\det(G))$ is a correction term (called the Fixman correction) that is needed to obtain the convergence to the constrained dynamics (1.2) (see [34, Sect. 3.2.3.4] and references therein). Thus, for ε small, the trajectory of the solution of (1.3) lies in the vicinity of the manifold $\mathcal{M}$. This penalization can also appear naturally when simulating Langevin dynamics with a stiff potential (see, for instance, [45, Sect. 5.1]). One is then interested in numerical schemes that are robust with respect to the parameter ε and that lie on the manifold $\mathcal{M}$ in the limit $\varepsilon \to 0$. In this paper, we study the following similar penalized dynamics in $\mathbb{R}^d$ to simulate trajectories in a vicinity of the manifold $\mathcal{M}$:

$$dX^\varepsilon(t) = f(X^\varepsilon(t))dt + \sigma dW(t) + \frac{\sigma^2}{4}\nabla\ln(\det(G))(X^\varepsilon(t))dt - \frac{1}{\varepsilon}(gG^{-1}\zeta)(X^\varepsilon(t))dt, \quad (1.4)$$

where $X^\varepsilon(0) = X_0$. It is a simpler version of (1.3) that also evolves in a vicinity of the manifold $\mathcal{M}$ in the limit $\varepsilon \to 0$. One result of this paper is the strong convergence of the solution X^ε of (1.4) to the solution X^0 of (1.2) if $X_0 \in \mathcal{M}$. We mention that in the deterministic setting, that is, when $\sigma = 0$, equation (1.4) is a singular perturbation problem, and it converges to a differential algebraic equation (DAE) of index two in the limit $\varepsilon \to 0$ (see [22, Chaps. VI-VII]). We propose in this article a method that is robust with respect to the parameter ε for solving equations of the form (1.4), and we leave the creation of robust integrators for solving (1.3) for future work for the sake of clarity.

There are different ways to approximate the solution of the dynamics (1.4). A strong approximation focuses on approximating the realization of a single trajectory of (1.4) for a given realization of the Wiener process W. A weak approximation approximates the average of functionals of the solution at a fixed time T, that is, quantities of the form $\mathbb{E}[\phi(X^\varepsilon(T))]$ for ϕ a smooth test function. In addition, under growth and smoothness assumptions on the vector fields in (1.4) (see, for instance, [23]), the dynamics (1.4) naturally satisfy an ergodicity property; that is, there exists a unique invariant measure $d\mu_\infty^\varepsilon$ in $\mathbb{R}^d$ that has a density ρ_∞^ε with respect to the Lebesgue measure, such that for all test functions ϕ,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \phi(X(t))dt = \int_{\mathbb{R}^d}\phi(x)d\mu_\infty^\varepsilon(x) \quad \text{almost surely.}$$

An approximation for the invariant measure focuses on approximating the average of a functional in the stationary state, that is, the quantity $\int_{\mathbb{R}^d}\phi(x)d\mu_\infty^\varepsilon(x)$. This is a computational

challenge when the dimension d is high, which is the case in the context of molecular dynamics where the dimension is proportional to the number of particles, as a standard quadrature formula becomes prohibitively expensive in high dimension. We emphasize that the invariant measure μ_∞^ε becomes singular with respect to the Lebesgue measure on $\mathbb{R}^d$ in the limit $\varepsilon \to 0$, and tends weakly as $\varepsilon \to 0$ to $d\mu_\infty^0$, a measure that is absolutely continuous to $d\sigma_{\mathcal{M}}$, the canonical measure on $\mathcal{M}$ induced by the Euclidean metric of $\mathbb{R}^d$. In this paper, we propose weak convergence results for a new uniformly accurate integrator for solving (1.4), and numerical experiments in the weak context and for the invariant measure, as we recall that a scheme of weak order r automatically has order $p \geqslant r$ for the invariant measure (see, for instance, [37]).

The most used discretization for solving (1.4) is the explicit Euler integrator in $\mathbb{R}^d$ (see [15, 33, 34, 35], for instance),

$$X_{n+1} = X_n + \sqrt{h}\sigma\xi + hf(X_n) + h\frac{\sigma^2}{4}\nabla\ln(\det(G))(X_n) - \frac{h}{\varepsilon}(gG^{-1}\zeta)(X_n). \qquad (1.5)$$

This integrator has weak order one of accuracy, but it faces some severe stepsize restriction due to its instability, typically of the form $h \ll \varepsilon$, in order to be accurate in the regime $\varepsilon \to 0$. Since the solution $X^\varepsilon(t)$ of (1.4) converges to the solution $X^0(t)$ of (1.2) when $\varepsilon \to 0$, one can use integrators for the limit equation (1.2) and apply them to solve the original problem (1.4) when ε is close to zero. Indeed, one can prove that the solution $X^\varepsilon(t)$ of (1.4) stays at distance $\mathcal{O}(\sqrt{\varepsilon})$ of the constrained solution $X^0(t)$ of (1.2) (see Theorem 2.3). Thus, if the timestep of the integrator is small enough and satisfies $\varepsilon \ll h$, then this integrator is consistent for solving (1.4). The alternative for the discretization of (1.2) on the manifold of the explicit Euler scheme (1.5) is the constrained Euler scheme

$$X_{n+1}^0 = X_n^0 + \sqrt{h}\sigma\xi + hf(X_n^0) + g(X_n^0)\lambda_{n+1}^0, \quad \zeta(X_{n+1}^0) = 0, \qquad (1.6)$$

where $\lambda_{n+1}^0 \in \mathbb{R}^q$ acts as a Lagrange multiplier, and is entirely determined by the constraint $\zeta(X_{n+1}^0) = 0$. This integrator has weak order one for solving (1.2) (see [34, Sect. 3.2.4]) and it lies on the manifold $\mathcal{M}$. It is a consistent approximation of (1.4) if ε is close to zero and $\varepsilon \ll h$. This integrator is, however, not appropriate for solving (1.4) if the size of ε is of the order of one, since the exact solution does not evolve in a neighborhood of the manifold in this regime. We mention a few other techniques to integrate numerically (1.4) or (1.2). In [28, 30], high order Runge-Kutta methods are proposed for sampling the invariant measure in $\mathbb{R}^d$ and on manifolds. The paper [36] presents a constrained integrator based on the RATTLE scheme (see [42, 5, 22]) in the context of the underdamped Langevin dynamics. Some of the previously cited discretizations can be combined with Metropolis-Hastings rejection procedures [39, 24], as done, for instance, in [20, 8, 35, 46, 36]. As for the Euler integrators (1.5)-(1.6), all of the previously mentioned methods are consistent provided that the parameter ε and the timestep satisfy $h \ll \varepsilon$ in the context of methods in $\mathbb{R}^d$, and satisfy $\varepsilon \ll h$ in the context of methods on the manifold $\mathcal{M}$. When applied in the regime where ε and h share the same order of magnitude, the accuracy of these methods quickly deteriorates, and they may face stability issues.

In past decades, different solutions were proposed for treating the loss of accuracy in the intermediate regime $h \sim \varepsilon$ in the context of multiscale problems with the help of uniformly accurate (UA) methods. These methods are capable of solving dynamics indexed by a possibly stiff parameter ε with an accuracy and a cost both independent of ε. A uniformly accurate method is automatically asymptotic-preserving (AP), that is, it converges in the two

regimes $\varepsilon \to 0$ and $\varepsilon \sim 1$, but the converse is not true in general. We refer the reader to the review [25] and references therein for examples of AP integrators for solving multiscale problems. We mention in particular the paper [41], which proposes a penalized Hamiltonian dynamics, and AP discretizations for solving it, and the paper [7] that gives an AP scheme for the approximation of a class of multiscale SDEs. In [17, 44, 29], trigonometric and multirevolution integrators are considered for solving highly oscillatory SDEs (see also the deterministic works [38, 9, 10, 14]). We mention the recent papers [11, 12, 13, 4] (see also the references therein) that introduce uniformly accurate methods for solving a variety of multiscale problems. There is a rich literature on AP and UA methods, but, to the best of our knowledge, the problem we study here and the techniques we consider are new. We propose in this paper a new consistent integrator with uniform accuracy and uniform cost for solving penalized Langevin dynamics, that is, a method for solving (1.4) whose accuracy and cost do not depend on the parameter ε.

The article is organized as follows. Section 2 is devoted to the presentation of the new integrator and to the main convergence results. In Section 3, we build a weak asymptotic expansion of the solution of (1.4) that is uniform in ε, and we use it for proving the uniform accuracy of our integrator. We compare in Section 4 the new integrator with the explicit Euler scheme in $\mathbb{R}^d$ (1.5) and the constrained Euler scheme (1.6) on $\mathcal{M}$ in numerical experiments on a torus and on the orthogonal group to confirm its order of convergence in the weak context and for sampling the invariant measure. Finally, we present some possible extensions and future work in Section 5.

2 Uniformly accurate integrator for penalized Langevin dynamics

In this section, we present the new uniformly accurate integrator and the main convergence results of this paper. The proofs are postponed to Section 3. Let us first lay down a few notations and assumptions. We assume in the rest of the article that $\mathcal{M} = \{x \in \mathbb{R}^d, \zeta(x) = 0\}$ is a compact and smooth manifold of codimension $q \geqslant 1$ embedded in $\mathbb{R}^d$, where the constraints are given by the smooth map $\zeta \colon \mathbb{R}^d \to \mathbb{R}^q$. We write $g = \nabla \zeta \colon \mathbb{R}^d \to \mathbb{R}^{d \times q}$ and we assume that the Gram matrix $G(x) = g^T(x) g(x) \in \mathbb{R}^{q \times q}$ is invertible for all x in $\mathcal{M}$. With these notations, the projection $\Pi_{\mathcal{M}}$ on the tangent bundle is given by $\Pi_{\mathcal{M}}(x) = I_d - G^{-1}(x) g(x) g^T(x)$. The test functions typically belong to a subspace of $\mathcal{C}_P^\infty(\mathbb{R}^d, \mathbb{R})$, the vector space of $\mathcal{C}^\infty$ functions $\phi(x)$ such that all partial derivatives up to any order have a polynomial growth of the form

$$\left| \phi^{(k)}(x) \right| \leqslant C(1 + |x|^K),$$

where the constants C and K are independent of $x \in \mathbb{R}^d$ (but can depend on k), and where we denote by $|x| = (x^T x)^{1/2}$ the Euclidean norm in $\mathbb{R}^d$. Similarly, we denote by $\mathcal{C}_P^p(\mathbb{R}^d, \mathbb{R})$ the space of $\mathcal{C}^p$ functions whose partial derivatives up to order p have polynomial growth. Letting $\varphi \colon \mathbb{R}^d \to \mathbb{R}^{d \times k}$, we use the following notations for differentials: for all vectors x, a^1, $\dots$, $a^m \in \mathbb{R}^d$, we denote

$$\varphi^{(m)}(x)(a^1, \dots, a^m) = \sum_{i_1, \dots, i_m = 1}^d \frac{\partial^m \varphi}{\partial x_{i_1} \dots \partial x_{i_m}}(x) \, a_{i_1}^1 \dots a_{i_m}^m.$$

For the sake of clarity, we will often drop the coefficient x, and if $m = 1$, we also use the notation φ' for the Jacobian matrix of φ. Moreover, for (e_i) the canonical basis of $\mathbb{R}^d$, we

write

$$\Delta\varphi(x) = \sum_{i=1}^{d} \varphi''(x)(e_i, e_i) \quad \text{and} \quad (\text{div}\,\varphi(x))_j = \sum_{i=1}^{d} \varphi'_{ij}(x)(e_i).$$

In the rest of the paper, we make the following assumption in the spirit of the regularity assumptions made in [16, Appx. C] and [40, Chap. 2].

Assumption 2.1. *The map f is bounded, is Lipschitz and lies in $\mathcal{C}_P^3$. The maps g and g' are bounded in $\mathbb{R}^d$, and there exists $c > 0$ such that, for x, $y \in \mathbb{R}^d$,*

$$|G_y(x)| \geqslant c(1 + |y|)^{-1}, \quad \text{where} \quad G_y(x) = \int_0^1 g^T(x + \tau y)d\tau g(x). \tag{2.1}$$

In addition, there exists a smooth change of coordinate

$$\psi\colon \mathbb{R}^d \to \mathbb{R}^d$$

$$x \mapsto \begin{pmatrix} \varphi(x) \\ \zeta(x) \end{pmatrix}$$

where $\varphi\colon \mathbb{R}^d \to \mathbb{R}^{d-q}$ satisfies $\varphi'g = 0$. The map ψ lies in $\mathcal{C}_P^5$ and is invertible, ψ' and ψ'' are Lipschitz and there exist two constants c, $C > 0$ such that $c \leqslant |\psi'(x)| \leqslant C$.

Remark 2.2. *Assumption 2.1 is almost the same as the one given in [16, Appx. C] to prove the strong convergence of the dynamics (1.3) to the constrained dynamics (1.2). The difference lies in the additional estimate (2.1) that replaces the weaker assumption that the Gram matrix $G(x) = G_0(x)$ is invertible on $\mathcal{M}$. We use the estimate (2.1) for obtaining a uniform expansion of the Lagrange multipliers in the new method and for proving that the new method evolves in a neighborhood of the manifold (see Lemma 3.8). Note that the existence of the change of coordinate ψ is always valid in a neighborhood of the smooth manifold $\mathcal{M}$. The same goes for the estimate (2.1) for x in a neighborhood of $\mathcal{M}$ and y in a ball centered on zero. Assumption 2.1 is valid in particular if $\mathcal{M}$ is a vector subspace of $\mathbb{R}^d$, but it is quite restrictive. It would be interesting to extend the results of this paper under simpler regularity assumptions made only on the manifold, as numerical experiments hint that the results presented in this paper still stand without global assumptions. This is matter for future work.*

Under Assumption 2.1, the problems (1.2) and (1.4) are well posed, and we obtain the strong convergence of the penalized dynamics (1.4) to the constrained dynamics (1.2).

Theorem 2.3. *Under Assumption 2.1, the solution X^ε of the penalized dynamics (1.4) converges strongly to X^0, the solution of the constrained dynamics (1.2); that is, for all $t \leqslant T$, there exists a constant $C > 0$ such that, for all $\varepsilon > 0$,*

$$\sup_{t \leqslant T} \mathbb{E}\Big[\big|X^\varepsilon(t) - X^0(t)\big|^2 \Big] \leqslant C\varepsilon.$$

Moreover, $\zeta(X^\varepsilon(t))$ satisfies

$$\sup_{t \leqslant T} \mathbb{E}\Big[|\zeta(X^\varepsilon(t))|^2 \Big] \leqslant C\varepsilon.$$

This result was first introduced in [16, Appx. C] for slightly different penalized dynamics. The proof is almost identical, but we present it in Appx. A for the sake of completeness. In the deterministic setting with $\sigma = 0$, the convergence to the manifold is of order 1 in ε instead of order $1/2$. We introduce the following additional assumption.

Assumption 2.4. *There exists $c > 0$ such that, for x, $y \in \mathbb{R}^d$,*

$$|G_y(x)| \geqslant c, \quad \text{where} \quad G_y(x) = \int_0^1 g^T(x + \tau y) d\tau g(x).$$

Assumption (2.4) is a stronger version of the inequality (2.1) and is in the spirit of the concept of admissible Lagrange multipliers [36]. It is always satisfied for x in a neighborhood of the manifold $\mathcal{M}$ and y in a ball centered on zero if we assume that the Gram matrix $G(x) = G_0(x)$ is invertible on $\mathcal{M}$. We emphasize that we do not need this assumption for proving the uniform accuracy property of the new method, but we use it for obtaining uniform estimates in the regime $\varepsilon \to 0$ and on the numerical implementation of the uniformly accurate method.

We introduce the new integrator for approximating the penalized dynamics (1.4) with cost and accuracy independent of the parameter ε, and a cost comparable to that of the constrained Euler scheme (1.6) in terms of the number of evaluations of the functions f, ζ, g, and g'.

New Method (Uniform discretization of penalized overdamped Langevin dynamics)

$X_0^\varepsilon = X_0 \in \mathcal{M}$
for $n \geqslant 0$ **do**

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \sqrt{h}\sigma\xi_n + hf(X_n^\varepsilon) + \frac{(1 - e^{-h/\varepsilon})^2}{2}(g'(gG^{-1}\zeta)G^{-1}\zeta)(X_n^\varepsilon)$$

$$+ \frac{\sigma^2\varepsilon}{8}(1 - e^{-2h/\varepsilon})\nabla\ln(\det(G))(X_n^\varepsilon) + g(X_n^\varepsilon)\lambda_{n+1}^\varepsilon, \qquad (2.2)$$

$$\zeta(X_{n+1}^\varepsilon) = e^{-h/\varepsilon}\zeta(X_n^\varepsilon) + \sigma\sqrt{\frac{\varepsilon}{2}(1 - e^{-2h/\varepsilon})}g^T(X_n^\varepsilon)\xi_n$$

$$+ \varepsilon(1 - e^{-h/\varepsilon})\big(g^Tf + \frac{\sigma^2}{4}g^T\nabla\ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g)\big)(X_n^\varepsilon)$$

$$+ \sigma^2\Big(\varepsilon(1 - e^{-h/\varepsilon}) - \sqrt{\frac{\varepsilon h}{2}(1 - e^{-2h/\varepsilon})}\Big)$$

$$\times \Big(\sum_{i=1}^d (g'(gG^{-1}g^Te_i))^T gG^{-1}g^Te_i - \sum_{i=1}^d (g'(e_i))^T gG^{-1}g^Te_i\Big)(X_n^\varepsilon).$$

end for

The new method works in a way similar to the constrained Euler integrator (1.6). Knowing the approximation X_n^ε of $X^\varepsilon(nh)$, we project a modified Euler step on a modified manifold defined by the constraint given in (2.2) (in place of $\zeta(X_{n+1}^\varepsilon) = 0$ for the constrained Euler integrator (1.6)). We project the modified step in the direction $g(X_n^\varepsilon)$ with the help of a Lagrange multiplier $\lambda_{n+1}^\varepsilon$. For the implementation of the method, one can use, for instance, a fixed point iteration or a Newton method at each step to find the solution $(X_{n+1}^\varepsilon, \lambda_{n+1}^\varepsilon)$ of the implicit system of equations (2.2).

In order for the discretization (2.2) to be well-defined, we use bounded random variables. The ξ_n are independent and bounded discrete random vectors that have the same moments as standard Gaussian random vectors up to order four, in the spirit of [40, Chap. 2], that is, for instance, that their components satisfy

$$\mathbb{P}(\xi_i = 0) = \frac{2}{3} \quad \text{and} \quad \mathbb{P}(\xi_i = \pm\sqrt{3}) = \frac{1}{6}, \quad i = 1, \ldots, d. \qquad (2.3)$$

Note that using truncated Gaussian random variables would also work.

Remark 2.5. *The new integrator is related to the popular idea of backward error analysis and modified equations for SDEs (see, for instance, [48, 1, 18, 26, 27]). The idea is to define a projection method (see [21, Sect. IV.4]) with a modified constraint in place of $\zeta(X_n) = 0$. Instead of evaluating the stiff term $\frac{h}{\varepsilon}gG^{-1}\zeta$ as in the Euler scheme (1.5), we project a modified step of the explicit Euler scheme in $\mathbb{R}^d$ on a manifold that is close to $\mathcal{M}$ when $\varepsilon \ll 1$ and whose constraint is given by a truncation of a uniform expansion of $\zeta(X^\varepsilon)$. When $\varepsilon \to \infty$, the expression of the constraint $\zeta(X_{n+1}^\varepsilon)$ in (2.2) tends to a truncated Taylor expansion in h around X_n^ε, while for $\varepsilon \to 0$, $\zeta(X_{n+1}^\varepsilon)$ tends to zero, which enforces that the integrator lies on $\mathcal{M}$. These intuitions will be made rigorous in Section 3.*

Remark 2.6. *In the context of a manifold $\mathcal{M}$ of codimension $q = 1$, the Gram matrix $G(x)$ and $g(x)^T e_i = g_i(x)$ are real numbers, so that $\nabla \ln(\det(G)) = 2G^{-1}g'(g)$ and*

$$\sum_{i=1}^d (g'(gG^{-1}g^T e_i))^T gG^{-1}g^T e_i = G^{-1}(g'(g))^T g = \sum_{i=1}^d (g'(e_i))^T gG^{-1}g^T e_i.$$

The discretization (2.2) thus reduces to

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \sqrt{h}\sigma\xi_n + hf(X_n^\varepsilon) + \frac{(1 - e^{-h/\varepsilon})^2}{2}(\zeta^2 G^{-2}g'(g))(X_n^\varepsilon)$$

$$+ \frac{\sigma^2\varepsilon}{4}(1 - e^{-2h/\varepsilon})(G^{-1}g'(g))(X_n^\varepsilon) + g(X_n^\varepsilon)\lambda_{n+1}^\varepsilon, \tag{2.4}$$

$$\zeta(X_{n+1}^\varepsilon) = e^{-h/\varepsilon}\zeta(X_n^\varepsilon) + \sigma\sqrt{\frac{\varepsilon}{2}(1 - e^{-2h/\varepsilon})}g^T(X_n^\varepsilon)\xi_n$$

$$+ \varepsilon(1 - e^{-h/\varepsilon})(g^T f + \frac{\sigma^2}{2}G^{-1}g^T g'(g) + \frac{\sigma^2}{2}\operatorname{div}(g))(X_n^\varepsilon).$$

We present in the rest of the section the uniform accuracy property of the discretization (2.2), and we show that the integrator converges to the constrained Euler scheme (1.6) when $\varepsilon \to 0$. The different convergence results are summarized by the following commutative diagram, where $T = Nh$ is fixed. Note that, as we present a convergence result in h that is uniform in ε, the two arrows for the convergence in h rely on the same Theorem 2.7.

$$
\begin{array}{ccc}
\text{integrator } X_N^\varepsilon \text{ in } \mathbb{R}^d & \xrightarrow[\text{(Thm. 2.9)}]{\varepsilon\to 0} & \text{integrator } X_N^0 \text{ on } \mathcal{M} \\
\Big\downarrow{\scriptstyle h\to 0}\,{\text{(Thm. 2.7)}} & & \Big\downarrow{\scriptstyle h\to 0}\,{\text{(Thm. 2.7)}} \\
\text{solution } X^\varepsilon(T) \text{ of (1.4)} & \xrightarrow[\text{(Thm. 2.3)}]{\varepsilon\to 0} & \text{solution } X^0(T) \text{ of (1.2)}
\end{array}
$$

We now state the main result of this work, that is, the uniform accuracy of the discretization (2.2) for approximating the solution of the penalized Langevin dynamics (1.4).

Theorem 2.7. *Under Assumption 2.1, the integrator (X_n^ε) given by (2.2) is a consistent uniformly accurate approximation of the solution $X^\varepsilon(t)$ of the penalized Langevin dynamics (1.4); that is, for a given test function $\phi \in \mathcal{C}_P^5$, there exist $h_0 > 0$, $C > 0$ such that for all $\varepsilon > 0$, $h \leqslant h_0$, the following estimate holds:*

$$|\mathbb{E}[\phi(X_n^\varepsilon)] - \mathbb{E}[\phi(X^\varepsilon(nh))]| \leqslant C\sqrt{h}, \quad n = 0, 1, \ldots, N, \quad Nh = T. \tag{2.5}$$

7

Remark 2.8. *Note that Theorem 2.7 states the uniform consistency, but not the uniform weak order one, as one could expect. The discretization (2.2) has weak order one if $\varepsilon = \varepsilon_0$ is fixed or in the limit $\varepsilon \to 0$, but an order reduction occurs in the intermediate regime $0 < \varepsilon < \varepsilon_0$, and the integrator only has weak order $1/2$ with respect to h in general. For the sake of simplicity, we leave the creation of uniformly accurate integrators of higher weak order for future works.*

We present the proof of Theorem 2.7 in Section 3. It relies on a weak expansion in h of the solution of (1.4) that is uniform in ε. One could directly use this uniform expansion as an explicit numerical integrator for solving (1.4). It would also yield a uniformly accurate scheme and would not require one to solve a fixed point problem. However, in the limit $\varepsilon \to 0$, this integrator would almost surely not stay on the manifold. The crucial geometric property that the integrator lies on the manifold when $\varepsilon \to 0$ is satisfied for the new method, as stated in the following result.

Theorem 2.9. *Under Assumption 2.1, the integrator (X_n^ε) in (2.2) converges to the Euler scheme on the manifold (1.6) when $\varepsilon \to 0$; that is, for h and N fixed such that $T = Nh$, there exists a constant $C_h > 0$ that depends on h but not on ε such that*

$$|\zeta(X_n^\varepsilon)| \leqslant C_h \sqrt{\varepsilon}, \quad n = 0, 1, \ldots, N.$$

In addition, if Assumption 2.4 is satisfied, then, for h_0 small enough, for $h \leqslant h_0$ fixed, there exists a constant $C_h > 0$ such that

$$\left| X_n^\varepsilon - X_n^0 \right| \leqslant C_h \sqrt{\varepsilon}, \quad n = 0, 1, \ldots, N, \quad Nh = T. \tag{2.6}$$

Remark 2.10. *In the deterministic context (i.e. when $\sigma = 0$), the uniform accuracy of the discretization (2.2) still holds, and the speed of convergence to the manifold $\mathcal{M}$ of both the exact solution and the integrator are in $\mathcal{O}(\varepsilon)$. To the best of our knowledge, the integrator given by (2.2) is the first integrator with the uniform accuracy property for solving the singular perturbation problem (1.4) with $\sigma = 0$. However, similar expansions that are uniform with respect to ε are presented in [22, Chaps. VI-VII] and references therein.*

Remark 2.11. *Another widely used scheme on the manifold is the Euler scheme with implicit projection direction,*

$$X_{n+1}^0 = X_n^0 + hf(X_n^0) + \sqrt{h}\sigma\xi_n + g(X_{n+1}^0)\lambda_{n+1}^0, \quad \zeta(X_{n+1}^0) = 0, \tag{2.7}$$

where the Lagrange multiplier λ_{n+1}^0 is determined by the constraint $\zeta(X_{n+1}^0) = 0$. The uniformly accurate discretization given in (2.2) can be modified so that it converges to the integrator (2.7) when $\varepsilon \to 0$. It suffices to replace the first line of (2.2)

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \sqrt{h}\sigma\xi_n + hf(X_n^\varepsilon) - \frac{(1 - e^{-h/\varepsilon})^2}{2}(g'(gG^{-1}\zeta)G^{-1}\zeta)(X_n^\varepsilon)$$

$$+ \frac{\sigma^2}{4}\left(\sqrt{2\varepsilon h(1 - e^{-2h/\varepsilon})} - \frac{\varepsilon}{2}(1 - e^{-2h/\varepsilon})\right)\nabla \ln(\det(G))(X_n^\varepsilon) + g(X_{n+1}^\varepsilon)\lambda_{n+1}^\varepsilon,$$

and to keep the same expansion for the constraint $\zeta(X_{n+1}^\varepsilon)$. The methodology for the uniform expansion of the integrator that we present in Section 3.2 extends to this context, so that the convergence results persist. Similarly, one could change the direction of projection $g(X_n^\varepsilon)$ into $g(Y_n^\varepsilon)$, where Y_n^ε is any consistent one-step approximation of X_n^ε, in the spirit of the class of projected Runge-Kutta methods presented in [30]. Finding a class of uniformly accurate discretizations that converge to a more general class of Runge-Kutta methods on the manifold $\mathcal{M}$ is matter for future works.

The uniform discretization (2.2) is implicit and requires one to solve a fixed point problem at each step with, for instance, a fixed point iteration or a Newton method. The following result, in the spirit of [22, Chap. VII] for deterministic DAEs and [30, Lemma 3.3] for the constrained dynamics (1.2), confirms that the associated implicit system is not stiff, that is, that its complexity does not depend on the stiff parameter ε.

Theorem 2.12. *Under Assumption 2.1, each step of the integrator (X_n^ε) given by (2.2) can be rewritten as a solution of a fixed point problem of the form*

$$X_{n+1}^\varepsilon = F_h^\varepsilon(X_{n+1}^\varepsilon),$$

where $F_h^\varepsilon \colon \mathbb{R}^d \to \mathbb{R}^d$ depends on X_n^ε, ξ_n, h, and ε. Moreover, if Assumption 2.4 is satisfied, then there exists $h_0 > 0$ independent of ε such that for all $h \leqslant h_0$, F_h^ε is a uniform contraction, that is, there exists a positive constant $L < 1$ independent of h and ε such that, for all $y_1, y_2 \in \mathbb{R}^d$,

$$|F_h^\varepsilon(y_2) - F_h^\varepsilon(y_1)| \leqslant L\,|y_2 - y_1|.$$

3 Weak convergence analysis

In this section, we present the uniform weak expansion and the stability properties of the solution $X^\varepsilon(t)$ to the penalized dynamic (1.4) and of the uniform integrator (X_n^ε) given in (2.2). We then use these results to prove Theorem 2.7, Theorem 2.9 and Theorem 2.12. Let us begin the analysis with a few technical lemmas and notations.

Lemma 3.1. *Let (X_n^ε) be given by (2.2), then, under Assumption 2.1, there exists a constant $C_0 > 0$ independent of $X_0 \in \mathcal{M}$, ε, and h such that, for all $n \geqslant 0$,*

$$(1 - e^{-h/\varepsilon})\,|\zeta(X_n^\varepsilon)| \leqslant C_0\sqrt{h}. \tag{3.1}$$

Proof. As f, ξ_n, g, and g' are bounded, we obtain from the definition of the integrator given by (2.2) that

$$|\zeta(X_{n+1}^\varepsilon)| \leqslant e^{-h/\varepsilon}|\zeta(X_n^\varepsilon)| + C\sqrt{\varepsilon(1 - e^{-2h/\varepsilon})},$$

where we used that the function $(1 - e^{-x})/x$ is bounded for $x > 0$. Thus, as $\zeta(X_0) = 0$,

$$|\zeta(X_n^\varepsilon)| \leqslant e^{-nh/\varepsilon}|\zeta(X_0)| + C\sqrt{\varepsilon(1 - e^{-2h/\varepsilon})} \sum_{k=0}^{n-1} e^{-kh/\varepsilon}$$

$$\leqslant C\frac{\sqrt{\varepsilon(1 - e^{-2h/\varepsilon})}}{1 - e^{-h/\varepsilon}} \leqslant C_0\frac{\sqrt{h}}{1 - e^{-h/\varepsilon}},$$

where the constant C_0 does not depend on X_0, ε, n, and h. This yields the estimate (3.1). $\quad\square$

The estimate (3.1) is a direct consequence of our choice of using bounded random variables in (2.2). We shall use this estimate extensively in the rest of this section. Thus, we denote by $\mathcal{M}_h^\varepsilon$ the set of vectors $x \in \mathbb{R}^d$ that satisfy the estimate

$$(1 - e^{-h/\varepsilon})\,|\zeta(x)| \leqslant C_0\sqrt{h}, \tag{3.2}$$

where C_0 is the constant given in Lemma 3.1. The set $\mathcal{M}_h^\varepsilon$ is a closed subset of $\mathbb{R}^d$ that contains $\mathcal{M}$. The numerical scheme given by (2.2) takes values in $\mathcal{M}_h^\varepsilon$. We mention that the

convergence results are still valid if the initial condition X_0 of (2.2) is chosen in $\mathcal{M}_h^\varepsilon$ instead of $\mathcal{M}$.

As we aim at writing uniform expansions, we introduce the convenient notation $R_h^\varepsilon(x)$ for any remainder that satisfies at least $|\mathbb{E}[R_h^\varepsilon(x)]| \leqslant Ch^{3/2}$, where C is independent of ε, h, and x.

The following result serves as a technical tool for simplifying the calculations in the uniform expansions of Subsection 3.1 and Subsection 3.2. It is proved with elementary computations.

Lemma 3.2. *The Fixman correction can be rewritten in the following way:*

$$\frac{\sigma^2}{4}\nabla\ln(\det(G)) = \frac{\sigma^2}{2}\sum_{i=1}^d g'(e_i)G^{-1}g^T e_i = \frac{\sigma^2}{2}\sum_{i=1}^d g'(gG^{-1}g^T e_i)G^{-1}g^T e_i,$$

where (e_i) is the canonical basis of $\mathbb{R}^d$.

Proof of Lemma 3.2. Using that $G = g^T g$ is a symmetric matrix, that $g = \nabla\zeta$ is a gradient (which implies $x^T g'(y) = y^T g'(x)$) and the standard properties of the trace operator Tr, we deduce that

$$\partial_j(\ln(\det(G))) = \text{Tr}(G^{-1}\partial_j G) = 2\,\text{Tr}(G^{-1}(g'(e_j))^T g) = 2\,\text{Tr}(g'(e_j)G^{-1}g^T)$$

$$= 2\sum_{i=1}^d e_i^T g'(e_j)G^{-1}g^T e_i = 2\sum_{i=1}^d e_j^T g'(e_i)G^{-1}g^T e_i,$$

that is, $\nabla\ln(\det(G)) = 2\sum_{i=1}^d g'(e_i)G^{-1}g^T e_i$. For the second equality, we have

$$\sum_{i=1}^d e_j^T g'(gG^{-1}g^T e_i)G^{-1}g^T e_i = \sum_{i=1}^d e_i^T gG^{-1}g^T g'(e_j)G^{-1}g^T e_i$$

$$= \text{Tr}(gG^{-1}g^T g'(e_j)G^{-1}g^T) = \text{Tr}(g'(e_j)G^{-1}g^T)$$

$$= \sum_{i=1}^d e_i^T g'(e_j)G^{-1}g^T e_i = \sum_{i=1}^d e_j^T g'(e_i)G^{-1}g^T e_i.$$

Hence we get the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.1 Uniform expansion of the exact solution

We consider the exact solution $X^\varepsilon(t)$ of the penalized Langevin dynamics (1.4), with the initial condition $X_0 = x$, that we assume is deterministic for simplicity. Then, X^ε satisfies the following expansion in h that is uniform with respect to ε.

Proposition 3.3. *Under Assumption 2.1, there exists $h_0 > 0$ such that for all $h \leqslant h_0$, if X^ε is the solution of the penalized Langevin dynamics (1.4) starting at $x \in \mathcal{M}_h^\varepsilon$, then, for all $\phi \in \mathcal{C}_P^3$, the following estimate holds:*

$$\left|\mathbb{E}[\phi(X^\varepsilon(h))] - \mathbb{E}[\phi(x + \sqrt{h}A_h^\varepsilon(x) + hB_h^\varepsilon(x))]\right| \leqslant C(1 + |x|^K)h^{3/2}, \qquad (3.3)$$

where C is independent of h and ε and where the functions A_h^ε and B_h^ε are given by

$$A_h^\varepsilon = \sigma\xi + \frac{e^{-h/\varepsilon} - 1}{\sqrt{h}}gG^{-1}\zeta + \sigma\left(\sqrt{\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon})} - 1\right)gG^{-1}g^T\xi,$$

$$B_h^\varepsilon = f + \left(\frac{\varepsilon}{h}(1 - e^{-h/\varepsilon}) - 1\right)gG^{-1}g^T f + \frac{\sigma^2}{2}\left(\frac{\varepsilon}{h}(1 - e^{-h/\varepsilon}) - 1\right)gG^{-1}\operatorname{div}(g)$$

$$+ \frac{\sigma^2\varepsilon}{8h}(1 - e^{-2h/\varepsilon})\nabla\ln(\det(G)) + \frac{\sigma^2\varepsilon}{8h}(1 - e^{-h/\varepsilon})^2 gG^{-1}g^T\nabla\ln(\det(G))$$

$$+ \frac{1}{2h}(e^{-h/\varepsilon} - 1)^2\Big(g'(gG^{-1}\zeta)G^{-1}\zeta - gG^{-1}g^T g'(gG^{-1}\zeta)G^{-1}\zeta$$

$$- gG^{-1}(g'(gG^{-1}\zeta))^T gG^{-1}\zeta\Big) + \sigma^2\Big(1 + \frac{\varepsilon}{h}(e^{-h/\varepsilon} - 1)\Big)\sum_{i=1}^d gG^{-1}(g'(e_i))^T gG^{-1}g^T e_i$$

$$+ \frac{\sigma^2}{4}\Big(\frac{\varepsilon}{h}(e^{-2h/\varepsilon} - 4e^{-h/\varepsilon} + 3) - 2\Big)\sum_{i=1}^d gG^{-1}(g'(gG^{-1}g^T e_i))^T gG^{-1}g^T e_i,$$

with ξ a discrete bounded random vector that satisfies (2.3), and where the functions A_h^ε and B_h^ε are bounded uniformly in ε and h on $\mathcal{M}_h^\varepsilon$.

Note that for $q = 1$, we obtain the first step of the uniformly accurate discretization (2.4) by gathering all the terms of the form gM with $M \in \mathbb{R}^q$ of the weak approximation given in Proposition 3.3 in a Lagrange multiplier $g\lambda_1^\varepsilon \in \mathbb{R}^q$ and by adding the truncated expansion of $\zeta(X^\varepsilon(h))$ as a constraint.

The proof of Proposition 3.3 relies on the change of coordinate ψ given in Assumption 2.1. Instead of discretizing directly the penalized dynamics (1.4), we first apply the change of coordinate ψ and we derive an expansion in time of $X^\varepsilon(t)$ that is uniform in the parameter ε. The following result is used for proving Proposition 3.3.

Lemma 3.4. *With the same notations and assumptions as in Proposition 3.3, the following estimates hold for all $h \leqslant h_0$ and all $x \in \mathcal{M}_h^\varepsilon$:*

$$\mathbb{E}[|X^\varepsilon(h) - x|^2]^{1/2} \leqslant C\sqrt{h}, \tag{3.4}$$

$$\mathbb{E}[|X^\varepsilon(h) - (x + \sqrt{h}\widehat{A}_h^\varepsilon(x))|^2]^{1/2} \leqslant Ch, \tag{3.5}$$

$$\left|\mathbb{E}[\psi(X^\varepsilon(h))] - \mathbb{E}[\psi(x + \sqrt{h}\widehat{A}_h^\varepsilon(x) + hB_h^\varepsilon(x))]\right| \leqslant Ch^{3/2}, \tag{3.6}$$

where C is independent of h and ε, B_h^ε is defined in Proposition 3.3, and $\widehat{A}_h^\varepsilon$ is given by

$$\widehat{A}_h^\varepsilon = \sigma\frac{W(h)}{\sqrt{h}} + \frac{e^{-h/\varepsilon} - 1}{\sqrt{h}}gG^{-1}\zeta + \frac{\sigma}{\sqrt{h}}gG^{-1}g^T\int_0^h (e^{(s-h)/\varepsilon} - 1)dW(s).$$

Proof of Proposition 3.3. The uniform bounds on $A_h^\varepsilon(X^\varepsilon(t))$ and $B_h^\varepsilon(X^\varepsilon(t))$ are obtained straightforwardly by using Assumption 2.1 and the fact that $x \in \mathcal{M}_h^\varepsilon$. We prove the local weak order one of the approximation $Y^\varepsilon(h) = x + \sqrt{h}\widehat{A}_h^\varepsilon(x) + hB_h^\varepsilon(x)$ given in Lemma 3.4. Let $\phi \in \mathcal{C}_P^3$ and $\widetilde{\phi} = \phi \circ \psi^{-1}$; then a Taylor expansion around x yields

$$\left|\mathbb{E}[\phi(X^\varepsilon(h)) - \phi(Y^\varepsilon(h))]\right| = \left|\mathbb{E}[\widetilde{\phi} \circ \psi(X^\varepsilon(h))] - \mathbb{E}[\widetilde{\phi} \circ \psi(Y^\varepsilon(h))]\right|$$

$$\leqslant \left|\mathbb{E}[\widetilde{\phi}'(\psi(x))(\psi(X^\varepsilon(h)) - \psi(Y^\varepsilon(h)))]\right|$$

$$+ \frac{1}{2}\Big|\mathbb{E}[\widetilde{\phi}''(\psi(x))(\psi(X^\varepsilon(h)) - \psi(x), \psi(X^\varepsilon(h)) - \psi(x))$$

$$- \widetilde{\phi}''(\psi(x))(\psi(Y^\varepsilon(h)) - \psi(x), \psi(Y^\varepsilon(h)) - \psi(x))]\Big|$$

$$+ C(1 + |x|^K)h^{3/2}$$

$$\leqslant \left| \widetilde{\phi}'(\psi(x)) \right| \left| \mathbb{E}[\psi(X^\varepsilon(h)) - \psi(Y^\varepsilon(h))] \right|$$

$$+ \frac{1}{2} \left| \widetilde{\phi}''(\psi(x)) \right| \mathbb{E}[|\psi(X^\varepsilon(h)) + \psi(Y^\varepsilon(h)) - 2\psi(x)|^2]^{1/2}$$

$$\cdot \mathbb{E}[|\psi(X^\varepsilon(h)) - \psi(Y^\varepsilon(h))|^2]^{1/2} + C(1 + |x|^K)h^{3/2}$$

where we used (3.4), Assumption 2.1, and the bilinearity of $\widetilde{\phi}''(\psi(x))$. With Lemma 3.4 and the regularity properties of $\widetilde{\phi}$ and ψ, we get

$$\left| \mathbb{E}[\phi(X^\varepsilon(h))] - \mathbb{E}[\phi(x + \sqrt{h}\widehat{A}_h^\varepsilon(x) + hB_h^\varepsilon(x))] \right| \leqslant C(1 + |x|^K)h^{3/2}. \tag{3.7}$$

In the spirit of [40, Chap. 2], we replace the random variable $\widehat{A}_h^\varepsilon(x)$ by the random variable $A_h^\varepsilon(x)$ that share the same expectation and covariance matrix. Indeed, a calculation gives

$$\text{Cov}(\widehat{A}_{h,i}^\varepsilon(x), \widehat{A}_{h,j}^\varepsilon(x)) = \sigma^2 \delta_{ij} + \frac{2\sigma^2}{h}(gG^{-1}g^T)_{ij} \int_0^h (e^{(s-h)/\varepsilon} - 1)ds$$

$$+ \frac{\sigma^2}{h} \sum_{k=1}^d (gG^{-1}g^T)_{ik}(gG^{-1}g^T)_{jk} \int_0^h (e^{(s-h)/\varepsilon} - 1)^2 ds$$

$$= \sigma^2 \delta_{ij} + \sigma^2(gG^{-1}g^T)_{ij}\left(\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon}) - 1 \right),$$

where we used that $g^T g = G$ and the Itô isometry. On the other hand, a similar calculation yields

$$\text{Cov}(A_{h,i}^\varepsilon(x), A_{h,j}^\varepsilon(x)) = \sigma^2 \delta_{ij} + 2\sigma^2\left(\sqrt{\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon})} - 1 \right)(gG^{-1}g^T)_{ij}$$

$$+ \sigma^2\left(\sqrt{\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon})} - 1 \right)^2 \sum_{k=1}^d (gG^{-1}g^T)_{ik}(gG^{-1}g^T)_{jk}$$

$$= \sigma^2 \delta_{ij} + \sigma^2(gG^{-1}g^T)_{ij}\left(\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon}) - 1 \right).$$

Replacing $\widehat{A}_h^\varepsilon(x)$ by $A_h^\varepsilon(x)$ in the weak expansion (3.7) gives the estimate (3.3). $\qquad \square$

The main ingredient of the proof of Lemma 3.4 is the decomposition of the terms of the expansion in a part that stays on the tangent space and a part of the form gM with $M \in \mathbb{R}^q$ that is orthogonal to the tangent space.

Proof of Lemma 3.4. As ψ^{-1} is Lipschitz, we have

$$\mathbb{E}[|X^\varepsilon(h) - x|^2]^{1/2} \leqslant C\mathbb{E}[|\psi(X^\varepsilon(h)) - \psi(x)|^2]^{1/2}$$

$$\leqslant C\mathbb{E}[|\varphi(X^\varepsilon(h)) - \varphi(x)|^2]^{1/2} + C\mathbb{E}[|\zeta(X^\varepsilon(h)) - \zeta(x)|^2]^{1/2}.$$

On the one hand, applying the Itô formula to $\varphi(X^\varepsilon)$ yields

$$\varphi(X^\varepsilon(h)) = \varphi(x) + \sigma \int_0^h \varphi'(X^\varepsilon(s))dW(s) \tag{3.8}$$

12

$$+ \int_0^h [\varphi' f + \frac{\sigma^2}{4}\varphi'\nabla \ln(\det(G)) + \frac{\sigma^2}{2}\Delta\varphi](X^\varepsilon(s))ds,$$

where the term in ε vanishes as $\varphi'g = 0$ (see Assumption 2.1). Assumption 2.1 allows us to write the uniform strong expansion

$$\mathbb{E}[|\varphi(X^\varepsilon(h)) - \varphi(x)|^2]^{1/2} \leqslant C\sqrt{h}.$$

On the other hand, for $\zeta(X^\varepsilon)$, we have

$$d\zeta(X^\varepsilon) = \sigma g^T(X^\varepsilon)dW + \left[g^T f + \frac{\sigma^2}{4}g^T\nabla \ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g) - \frac{1}{\varepsilon}\zeta\right](X^\varepsilon)dt.$$

With the variation of constants formula, it rewrites into

$$\zeta(X^\varepsilon(h)) = e^{-h/\varepsilon}\zeta(x) + \sigma\int_0^h e^{(s-h)/\varepsilon}g^T(X^\varepsilon(s))dW(s) \tag{3.9}$$

$$+ \int_0^h e^{(s-h)/\varepsilon}\left[g^T f + \frac{\sigma^2}{4}g^T\nabla\ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g)\right](X^\varepsilon(s))ds.$$

As the integrands in (3.9) are bounded (using Assumption 2.1), we get

$$\mathbb{E}[|\zeta(X^\varepsilon(h)) - \zeta(x)|^2]^{1/2} \leqslant C((e^{-h/\varepsilon} - 1)^2\,|\zeta(x)|^2 + h)^{1/2} \leqslant C\sqrt{h},$$

where we used that $x \in \mathcal{M}_h^\varepsilon$. We thus get the desired estimate (3.4). The estimate (3.5) is obtained with the same arguments by keeping track of the terms of size $\mathcal{O}(\sqrt{h})$ in the expansions.

We now prove the weak estimate (3.6). We denote for simplicity $Y^\varepsilon(h) = x + \sqrt{h}\widehat{A}_h^\varepsilon(x) + hB_h^\varepsilon(x)$. Let us first look at the approximation of $\varphi(X^\varepsilon(h))$. On the one hand, applying the Itô formula to $\varphi(X^\varepsilon(t))$ gives

$$\varphi(X^\varepsilon(h)) = \varphi(x) + h\varphi' f(x) + h\frac{\sigma^2}{4}\varphi'\nabla\ln(\det(G))(x) + h\frac{\sigma^2}{2}\Delta\varphi(x) + R_h^\varepsilon(x),$$

where we used (3.4) and we put in $R_h^\varepsilon(x)$ all the terms that are zero in average. On the other hand, an expansion in h of $\varphi(Y^\varepsilon(h))$ yields

$$\varphi(Y^\varepsilon(h)) = \varphi + \sqrt{h}\varphi'\widehat{A}_h^\varepsilon + h\left[\varphi'B_h^\varepsilon + \frac{1}{2}\varphi''(\widehat{A}_h^\varepsilon, \widehat{A}_h^\varepsilon)\right] + R_h^\varepsilon$$

$$= \varphi + h\varphi' f + \frac{\sigma^2\varepsilon}{8}(1 - e^{-2h/\varepsilon})\varphi'\nabla\ln(\det(G)) + \frac{1}{2}(e^{-h/\varepsilon} - 1)^2\varphi'g'(gG^{-1}\zeta)G^{-1}\zeta$$

$$+ \frac{\sigma^2}{2}\varphi''(W(h), W(h)) + \frac{1}{2}(e^{-h/\varepsilon} - 1)^2\varphi''(gG^{-1}\zeta, gG^{-1}\zeta)$$

$$+ \frac{\sigma^2}{2}\int_0^h\int_0^h (e^{(s-h)/\varepsilon} - 1)(e^{(u-h)/\varepsilon} - 1)\varphi''(gG^{-1}g^T dW(s), gG^{-1}g^T dW(u))$$

$$+ \sigma^2\int_0^h (e^{(s-h)/\varepsilon} - 1)\varphi''(W(h), gG^{-1}g^T dW(s)) + R_h^\varepsilon,$$

where we use that $\varphi'g = 0$, we omit the dependence in x for conciseness, and we put in $R_h^\varepsilon(x)$ all the terms that are zero in average. We now replace the random terms by their expectation,

$$\varphi(Y^\varepsilon(h)) = \varphi + h\varphi' f + \frac{\sigma^2\varepsilon}{8}(1 - e^{-2h/\varepsilon})\varphi'\nabla\ln(\det(G)) + h\frac{\sigma^2}{2}\Delta\varphi$$

13

$$+ \frac{1}{2}(e^{-h/\varepsilon} - 1)^2(\varphi'(g'(gG^{-1}\zeta)G^{-1}\zeta) + \varphi''(gG^{-1}\zeta, gG^{-1}\zeta))$$

$$+ \frac{\sigma^2}{2}\int_0^h (e^{(s-h)/\varepsilon} - 1)^2 ds \sum_{i=1}^d \varphi''(gG^{-1}g^T e_i, gG^{-1}g^T e_i)$$

$$+ \sigma^2 \int_0^h (e^{(s-h)/\varepsilon} - 1) ds \sum_{i=1}^d \varphi''(e_i, gG^{-1}g^T e_i) + R_h^\varepsilon.$$

Letting $M \in \mathbb{R}^q$, we differentiate the equality $\varphi'(gM) = 0$. We obtain that for any $M \in \mathbb{R}^q$ and $v \in \mathbb{R}^d$ we have $\varphi'(g'(v)M) + \varphi''(gM, v) = 0$. We deduce that

$$\varphi'(g'(gG^{-1}\zeta)G^{-1}\zeta) + \varphi''(gG^{-1}\zeta, gG^{-1}\zeta) = 0.$$

Applying Lemma 3.2, we get by a direct calculation that

$$\varphi(Y^\varepsilon(h)) = \varphi(x) + h\varphi' f(x) + h\frac{\sigma^2}{4}\varphi'\nabla \ln(\det(G))(x) + h\frac{\sigma^2}{2}\Delta\varphi(x) + R_h^\varepsilon(x),$$

which gives the desired estimate

$$\left|\mathbb{E}[\varphi(X^\varepsilon(h))] - \mathbb{E}[\varphi(x + \sqrt{h}\widehat{A}_h^\varepsilon(x) + hB_h^\varepsilon(x))]\right| \leqslant Ch^{3/2}. \tag{3.10}$$

For the one-step approximation of $\zeta(X^\varepsilon(h))$, the Itô formula and the variation of constants formula yield

$$\zeta(X^\varepsilon(h)) = e^{-h/\varepsilon}\zeta(x) + \varepsilon(1 - e^{-h/\varepsilon})(g^T f + \frac{\sigma^2}{4}g^T\nabla\ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g))(x) + R_h^\varepsilon(x).$$

For $\zeta(Y^\varepsilon(h))$, using $g^T g = G$, we get with the same arguments as for $\varphi(Y^\varepsilon(h))$ that

$$\zeta(Y^\varepsilon(h)) = \zeta + \sqrt{h}g^T\widehat{A}_h^\varepsilon + h\left[g^T B_h^\varepsilon + \frac{1}{2}(g'(\widehat{A}_h^\varepsilon))^T\widehat{A}_h^\varepsilon\right] + R_h^\varepsilon$$

$$= e^{-h/\varepsilon}\zeta + \varepsilon(1 - e^{-h/\varepsilon})g^T f + \frac{\sigma^2}{2}\left(\varepsilon(1 - e^{-h/\varepsilon}) - h\right)\operatorname{div}(g) + \frac{\sigma^2}{2}(g'(W(h)))^T W(h)$$

$$+ \frac{\sigma^2 \varepsilon}{4}(1 - e^{-h/\varepsilon})^2 g^T\nabla\ln(\det(G)) + \sigma^2\left(h + \varepsilon(e^{-h/\varepsilon} - 1)\right)\sum_{i=1}^d (g'(e_i))^T gG^{-1}g^T e_i$$

$$+ \frac{\sigma^2}{4}\left(\varepsilon(e^{-2h/\varepsilon} - 4e^{-h/\varepsilon} + 3) - 2h\right)\sum_{i=1}^d (g'(gG^{-1}g^T e_i))^T gG^{-1}g^T e_i$$

$$+ \frac{\sigma^2}{2}\int_0^h\int_0^h (e^{(s-h)/\varepsilon} - 1)(e^{(u-h)/\varepsilon} - 1)(g'(gG^{-1}g^T dW(u)))^T gG^{-1}g^T dW(s)$$

$$+ \sigma^2\int_0^h\int_0^h (e^{(s-h)/\varepsilon} - 1)(g'(W(h)))^T gG^{-1}g^T dW(s) + R_h^\varepsilon.$$

We now replace the stochastic integrals by their expectations (putting the remainders in R_h^ε), and we use Lemma 3.2 to simplify the expansion. It yields

$$\zeta(Y^\varepsilon(h)) = e^{-h/\varepsilon}\zeta + \varepsilon(1 - e^{-h/\varepsilon})(g^T f + \frac{\sigma^2}{4}g^T\nabla\ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g)) + R_h^\varepsilon,$$

which implies

$$\left|\mathbb{E}[\zeta(X^\varepsilon(h))] - \mathbb{E}[\zeta(x + \sqrt{h}\widehat{A}_h^\varepsilon(x) + hB_h^\varepsilon(x))]\right| \leqslant Ch^{3/2}. \tag{3.11}$$

Combining the inequalities (3.10) and (3.11) gives the desired weak estimate (3.6). $\qquad \square$

14

To end this subsection, we recall the growth properties of X^ε that will be of use in Subsection 3.3. For this particular result, we add the dependency in the initial condition x of the exact solution of (1.4) with the notation $X^\varepsilon(t, x)$.

Lemma 3.5. *Under Assumption 2.1, for $\phi \in \mathcal{C}_P^5$ and $t \leqslant T$ fixed, the function $\widetilde{\phi}^\varepsilon(x) = \mathbb{E}[\phi(X^\varepsilon(t, x))]$ lies in $\mathcal{C}_P^3$ with constants independent of t and ε.*

Proof. The standard theory (see, for instance, the textbook [40]) gives $\widetilde{\phi}^\varepsilon \in \mathcal{C}^3$. With the regularity assumptions on ψ and its derivatives, it is sufficient to prove that $\mathbb{E}[\phi(Y^\varepsilon(t, y))]$ is in $\mathcal{C}_P^3$, where $Y^\varepsilon(t, y) = \psi(X^\varepsilon(t, \psi^{-1}(x)))$ (replacing ϕ by $\phi \circ \psi^{-1}$). We recall from the proof of Lemma 3.4 that $\varphi(X^\varepsilon(t, x))$ satisfies the integral formulation (3.8) and that $\zeta(X^\varepsilon(t, x))$ satisfies (3.9). Putting together these equations, we deduce that $Y^\varepsilon(t, y)$ satisfies an equation of the form

$$Y^\varepsilon(t, y) = A^\varepsilon(t)y + \int_0^t A^\varepsilon(t - s)F(Y^\varepsilon(s, y))ds + \int_0^t A^\varepsilon(t - s)G(Y^\varepsilon(s, y))dW(s),$$

where F and G are in $\mathcal{C}_P^3$ and do not depend on ε, and

$$A^\varepsilon(t) = \begin{pmatrix} I_{d-q} & 0 \\ 0 & e^{-t/\varepsilon}I_q \end{pmatrix}.$$

As $|A^\varepsilon(t)| \leqslant C$, the process $Y^\varepsilon(t, y)$ satisfies $\mathbb{E}[|Y^\varepsilon(t, y)|^{2p}] \leqslant C(1 + |y|^K)$, with C and K independent of ε. Thus we have

$$\mathbb{E}[\phi(Y^\varepsilon(t, y))] \leqslant C(1 + \mathbb{E}[|Y^\varepsilon(t, y)|^K]) \leqslant C(1 + \mathbb{E}[|y|^K]).$$

For the derivatives, we recall from [19] that $Z^\varepsilon(t, y) = \partial_y Y^\varepsilon(t, y)$ satisfies the equation

$$Z^\varepsilon(t, y) = A^\varepsilon(t)I_d + \int_0^t A^\varepsilon(t - s)F'(Y^\varepsilon(s, y))Z^\varepsilon(t, y)ds$$

$$+ \int_0^t A^\varepsilon(t - s)G'(Y^\varepsilon(s, y))Z^\varepsilon(t, y)dW(s).$$

Applying the same arguments as for $Y^\varepsilon(t, y)$ yields that $Z^\varepsilon(t, y)$ has bounded moments of all order and that $\mathbb{E}[\phi(Z^\varepsilon(t, y))] \leqslant C(1 + \mathbb{E}[|y|^K])$. The same methodology extends to $\partial_y^2 Y^\varepsilon(t, y)$ and $\partial_y^3 Y^\varepsilon(t, y)$. $\qquad\square$

3.2 Uniform expansion and bounded moments of the numerical solution

In this subsection, we show that the integrator given by (2.2) has bounded moments of all order, that it lies on the manifold $\mathcal{M}$ in the limit $\varepsilon \to 0$, and that it satisfies the same local weak uniform expansion as the exact solution of (1.4) (see Proposition 3.3).

First, the integrator given by (2.2) satisfies the following bounded moments property.

Proposition 3.6. *Under Assumption 2.1, (X_n^ε) has bounded moments of any order along time; i.e., for all timestep $h \leqslant h_0$ small enough such that $Nh = T$ is fixed, for all integer $m \geqslant 0$,*

$$\sup_{n \leqslant N} \mathbb{E}[|X_n^\varepsilon|^{2m}] \leqslant C_m,$$

where the constant $C > 0$ is independent of ε and h.

The integrator (X_n^ε) given in (2.2) also satisfies a uniform local expansion that is similar to its continuous counterpart presented in Proposition 3.3.

Proposition 3.7. *Under Assumption 2.1, there exists $h_0 > 0$ such that for all $h \leqslant h_0$, if (X_n^ε) is the numerical discretization given by (2.2) beginning at $x \in \mathcal{M}_h^\varepsilon$ (assumed deterministic for simplicity), then, for all test functions $\phi \in \mathcal{C}_P^3$, the following estimate holds:*

$$\left| \mathbb{E}[\phi(X_1^\varepsilon)] - \mathbb{E}[\phi(x + \sqrt{h} A_h^\varepsilon(x) + h B_h^\varepsilon(x))] \right| \leqslant C(1 + |x|^K) h^{3/2}, \tag{3.12}$$

where C is independent of h and ε, and A_h^ε and B_h^ε are the functions given in Proposition 3.3.

To prove Proposition 3.6 and Proposition 3.7, we rely on the following lemma, whose proof is postponed to the end of this subsection. We emphasize that an inequality of the form $|\zeta(x)| \leqslant C$ does not imply in general that x stays close to $\mathcal{M}$. That is why we rely in Lemma 3.8 on an estimate of the Lagrange multipliers (using the inequality (2.1)). This estimate ensures that the method evolves in a neighborhood of the manifold.

Lemma 3.8. *Under Assumption 2.1 and if $x \in \mathcal{M}_h^\varepsilon$, there exists $h_0 > 0$ such that, for all timestep $h \leqslant h_0$, the one-step approximation X_1^ε and the Lagrange multiplier λ_1^ε in the discretization (2.2) satisfy*

$$|X_1^\varepsilon - x| \leqslant C\sqrt{h}, \qquad \lambda_1^\varepsilon = \sqrt{h} G^{-1}(x) \lambda_{1,(1/2)}^\varepsilon + h G^{-1}(x) \lambda_{1,(1)}^\varepsilon + R_h^\varepsilon(x),$$

where $|\lambda_{1,(1/2)}^\varepsilon| \leqslant C$, $|\lambda_{1,(1)}^\varepsilon| \leqslant C$, and $|R_h^\varepsilon(x)| \leqslant C h^{3/2}$ with C independent of ε, h and x.

For proving Proposition 3.6, we apply the change of variable ψ and we adapt the standard methodology presented in [40, Lemma 1.1.6 & Lemma 2.2.2].

Proof of Proposition 3.6. We derive from (2.2) that

$$\left| \mathbb{E}[\zeta(X_{n+1}^\varepsilon) - e^{-h/\varepsilon} \zeta(X_n^\varepsilon) | X_n^\varepsilon] \right| \leqslant Ch,$$

$$\left| \zeta(X_{n+1}^\varepsilon) - e^{-h/\varepsilon} \zeta(X_n^\varepsilon) \right| \leqslant C\sqrt{h}.$$

We prove that $\zeta(X_n)$ has bounded moments by induction on n. The binomial formula yields

$$\begin{aligned}
\mathbb{E}[|\zeta(X_{n+1}^\varepsilon)|^{2m}] &= \mathbb{E}\Big[\Big| e^{-h/\varepsilon} \zeta(X_n^\varepsilon) + \zeta(X_{n+1}^\varepsilon) - e^{-h/\varepsilon} \zeta(X_n^\varepsilon) \Big|^{2m} \Big] \\
&\leqslant e^{-2mh/\varepsilon} \mathbb{E}[|\zeta(X_n^\varepsilon)|^{2m}] \\
&\quad + C \mathbb{E}[|\zeta(X_n^\varepsilon)|^{2m-1} \Big| \mathbb{E}[\zeta(X_{n+1}^\varepsilon) - e^{-h/\varepsilon} \zeta(X_n^\varepsilon) | X_n^\varepsilon] \Big|] \\
&\quad + C \sum_{k=2}^{2m} \mathbb{E}[|\zeta(X_n^\varepsilon)|^{2m-k} \Big| \zeta(X_{n+1}^\varepsilon) - e^{-h/\varepsilon} \zeta(X_n^\varepsilon) \Big|^k] \\
&\leqslant \mathbb{E}[|\zeta(X_n^\varepsilon)|^{2m}] + C(1 + \mathbb{E}[|\zeta(X_n^\varepsilon)|^{2m}]) h.
\end{aligned}$$

Following [40, Lemma 1.1.6], as $X_0 \in \mathcal{M}$ is bounded, it implies that $\mathbb{E}[|\zeta(X_n)|^{2m}]$ is bounded uniformly in $n = 0, \dots, N$ and ε.

Using Lemma 3.8 and the equality $\varphi' g = 0$, a direct calculation gives

$$\left| \mathbb{E}[\varphi(X_{n+1}^\varepsilon) - \varphi(X_n^\varepsilon) | X_n^\varepsilon] \right| \leqslant Ch, \tag{3.13}$$

16

$$\left|\varphi(X^{\varepsilon}_{n+1}) - \varphi(X^{\varepsilon}_n)\right| \leqslant C\sqrt{h}. \tag{3.14}$$

Following the same methodology as for $\mathbb{E}[|\zeta(X_n)|^{2m}]$, the estimates (3.13)-(3.14) imply that the quantity $\mathbb{E}[|\varphi(X_n)|^{2m}]$ is bounded uniformly in $n = 0, \dots, N$ and ε. Then, as ψ^{-1} is Lipschitz, we have

$$\mathbb{E}[|X^{\varepsilon}_n|^{2m}] \leqslant C(1 + \mathbb{E}[|\psi(X^{\varepsilon}_n)|^{2m}]) \leqslant C(1 + \mathbb{E}[|\varphi(X^{\varepsilon}_n)|^{2m}] + \mathbb{E}[|\zeta(X^{\varepsilon}_n)|^{2m}]) \leqslant C.$$

Hence we get the result. $\qquad\square$

We obtain the uniform expansion of the numerical solution by writing explicitly a uniform expansion of the Lagrange multiplier λ^{ε}_1, in the spirit of [34, Lemma 3.25].

Proof of Proposition 3.7. Using Lemma 3.8 and Assumption 2.1, we obtain

$$X^{\varepsilon}_1 = x + \sqrt{h}\Big[\sigma\xi + (gG^{-1})(x)\lambda^{\varepsilon}_{1,(1/2)}\Big] + R^{\varepsilon}_h(x),$$

where the remainder satisfies $|R^{\varepsilon}_h(x)| \leqslant Ch$. The constraint is then given by

$$\zeta(X^{\varepsilon}_1) = \zeta(x) + \sqrt{h}\Big[\sigma g^T(x)\xi + \lambda^{\varepsilon}_{1,(1/2)}\Big] + R^{\varepsilon}_h(x).$$

On the other hand, we get from the definition of the integrator (2.2) that

$$\zeta(X^{\varepsilon}_1) = \zeta(x) + \sqrt{h}\Big[\frac{e^{-h/\varepsilon} - 1}{\sqrt{h}}\zeta(x) + \sigma\sqrt{\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon})}g^T(x)\xi\Big] + R^{\varepsilon}_h(x).$$

By identifying the two terms in $\sqrt{h}$ in the expansions of $\zeta(X^{\varepsilon}_1)$, we deduce the value of $\lambda^{\varepsilon}_{1,(1/2)}$, that is,

$$\lambda^{\varepsilon}_{1,(1/2)} = \frac{e^{-h/\varepsilon} - 1}{\sqrt{h}}\zeta + \sigma\Big(\sqrt{\frac{\varepsilon}{2h}(1 - e^{-2h/\varepsilon})} - 1\Big)g^T\xi.$$

The expression of $\lambda^{\varepsilon}_{1,(1/2)}$ and Lemma 3.8 give

$$X^{\varepsilon}_1 = x + \sqrt{h}A^{\varepsilon}_h(x) + h\Big[f(x) + \frac{(1 - e^{-h/\varepsilon})^2}{2h}(g'(gG^{-1}\zeta)G^{-1}\zeta)(x)$$
$$+ \frac{\sigma^2\varepsilon}{8h}(1 - e^{-2h/\varepsilon})\nabla\ln(\det(G))(x) + (gG^{-1})(x)\lambda^{\varepsilon}_{1,(1)}\Big] + R^{\varepsilon}_h(x), \tag{3.15}$$

where $|R^{\varepsilon}_h(x)| \leqslant Ch^{3/2}$. We then compute the expansion of $\zeta(X^{\varepsilon}_1)$, and we compare it with the definition of the integrator (2.2) to obtain the expression of $\lambda^{\varepsilon}_{1,(1)}$. Inserting this expression in (3.15) gives

$$X^{\varepsilon}_1 = x + \sqrt{h}A^{\varepsilon}_h(x) + hB^{\varepsilon}_h(x) + R^{\varepsilon}_h(x),$$

where the remainder satisfies $|\mathbb{E}[R^{\varepsilon}_h(x)]| \leqslant Ch^{3/2}$ and $|R^{\varepsilon}_h(x)| \leqslant Ch$. A Taylor expansion of $\phi(X^{\varepsilon}_1)$ around $x + \sqrt{h}A^{\varepsilon}_h(x) + hB^{\varepsilon}_h(x)$ yields the estimate (3.12). $\qquad\square$

The proof of Lemma 3.8 mainly relies on the estimates (2.1) and (3.2). We refer the reader to [30, Lemma 3.3] and [22, Chap. VII] for similar proofs where explicit expressions of Lagrange multipliers are derived.

Proof of Lemma 3.8. Let $x \in \mathcal{M}_h^\varepsilon$. For brevity, we rewrite the discretization (2.2) as

$$X_1^\varepsilon = x + \sqrt{h}Y_{(1/2)}^\varepsilon(x) + hY_{(1)}^\varepsilon(x) + g(x)\lambda_1^\varepsilon,$$
$$\zeta(X_1^\varepsilon) = \zeta(x) + \sqrt{h}\zeta_{(1/2)}^\varepsilon(x) + h\zeta_{(1)}^\varepsilon(x),$$

where the functions $Y_{(1/2)}^\varepsilon$, $Y_{(1)}^\varepsilon$, $\zeta_{(1/2)}^\varepsilon$, and $\zeta_{(1)}^\varepsilon$ are given by

$$Y_{(1/2)}^\varepsilon = \sigma\xi,$$
$$Y_{(1)}^\varepsilon = f + \frac{(1-e^{-h/\varepsilon})^2}{2h}(g'(gG^{-1}\zeta)G^{-1}\zeta) + \frac{\sigma^2\varepsilon}{8h}(1-e^{-2h/\varepsilon})\nabla\ln(\det(G)),$$
$$\zeta_{(1/2)}^\varepsilon = \frac{e^{-h/\varepsilon}-1}{\sqrt{h}}\zeta + \sigma\sqrt{\frac{\varepsilon}{2h}(1-e^{-2h/\varepsilon})}g^T\xi,$$
$$\zeta_{(1)}^\varepsilon = \frac{\varepsilon}{h}(1-e^{-h/\varepsilon})\left(g^Tf + \frac{\sigma^2}{4}g^T\nabla\ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g)\right)$$
$$+ \sigma^2\left(\frac{\varepsilon}{h}(1-e^{-h/\varepsilon}) - \sqrt{\frac{\varepsilon}{2h}(1-e^{-2h/\varepsilon})}\right)\sum_{i=1}^{d}((g'(gG^{-1}g^Te_i))^TgG^{-1}g^Te_i)$$
$$- \sigma^2\left(\frac{\varepsilon}{h}(1-e^{-h/\varepsilon}) - \sqrt{\frac{\varepsilon}{2h}(1-e^{-2h/\varepsilon})}\right)\sum_{i=1}^{d}((g'(e_i))^TgG^{-1}g^Te_i).$$

Using Assumption 2.1 and the estimate (3.2), the following uniform estimates hold:

$$\left|Y_{(i)}^\varepsilon(x)\right| \leqslant C, \quad \left|\zeta_{(i)}^\varepsilon(x)\right| \leqslant C, \quad i \in \{1/2, 1\}.$$

The fundamental theorem of calculus yields

$$\zeta(X_1^\varepsilon) - \zeta(x) = \int_0^1 g^T(x + \tau(X_1^\varepsilon - x))d\tau(X_1^\varepsilon - x) = \sqrt{h}\zeta_{(1/2)}^\varepsilon(x) + h\zeta_{(1)}^\varepsilon(x).$$

Substituting $X_1^\varepsilon - x$ then gives

$$\int_0^1 g^T(x + \tau(X_1^\varepsilon - x))d\tau(\sqrt{h}Y_{(1/2)}^\varepsilon(x) + hY_{(1)}^\varepsilon(x) + g(x)\lambda_1^\varepsilon) = \sqrt{h}\zeta_{(1/2)}^\varepsilon(x) + h\zeta_{(1)}^\varepsilon(x).$$

Using Assumption 2.1, we get the following explicit expression of λ_1^ε:

$$\lambda_1^\varepsilon = \sqrt{h}G_{X_1^\varepsilon-x}^{-1}(x)\left(\zeta_{(1/2)}^\varepsilon(x) - \int_0^1 g^T(x + \tau(X_1^\varepsilon - x))d\tau Y_{(1/2)}^\varepsilon(x)\right) \qquad (3.16)$$
$$+ hG_{X_1^\varepsilon-x}^{-1}(x)\left(\zeta_{(1)}^\varepsilon(x) - \int_0^1 g^T(x + \tau(X_1^\varepsilon - x))d\tau Y_{(1)}^\varepsilon(x)\right).$$

Then, the growth assumption (2.1) on $G_y^{-1}(x)$ allows us to write

$$|\lambda_1^\varepsilon| \leqslant C\sqrt{h}(1 + |X_1^\varepsilon - x|) \quad \text{and} \quad |X_1^\varepsilon - x| \leqslant C\sqrt{h}(1 + |X_1^\varepsilon - x|).$$

Hence, for $h \leqslant h_0$ small enough, we deduce that $|X_1^\varepsilon - x| \leqslant C\sqrt{h}$ and

$$\lambda_1^\varepsilon = \sqrt{h}G^{-1}(x)(\zeta_{(1/2)}^\varepsilon - g^TY_{(1/2)}^\varepsilon)(x) + R_h^\varepsilon, \qquad (3.17)$$

where $|R_h^\varepsilon| \leqslant Ch$. For the term of size $\mathcal{O}(h)$, we first deduce from (3.17) that

$$\left|X_1^\varepsilon - x - \sqrt{h}(Y_{(1/2)}^\varepsilon + gG^{-1}\zeta_{(1/2)}^\varepsilon - gG^{-1}g^TY_{(1/2)}^\varepsilon)(x)\right| \leqslant Ch.$$

By using this estimate in (3.16), a Taylor expansion yields the desired expansion of λ_1^ε. $\qquad\square$

3.3 Proofs of the convergence theorems

Now that we have the local uniform expansion of the exact solution and the numerical scheme, as well as the stability property of Proposition 3.6, we are able to prove the main convergence theorems.

Proof of Theorem 2.7. We derive the global weak consistency (2.5) with techniques similar to the ones presented in [40, Chap. 2]. We denote by $X^\varepsilon(t,x)$ the solution of the penalized dynamics with initial condition x and $X_n^\varepsilon(x)$ the numerical solution with initial condition x. For $x \in \mathcal{M}_h^\varepsilon$, Proposition 3.3 and Proposition 3.7 yield

$$\left|\mathbb{E}[\phi(X^\varepsilon(h,x)) - \phi(X_1^\varepsilon(x))|x]\right| \leqslant C(1 + |x|^K)h^{3/2},$$

where $\phi \in \mathcal{C}_P^3$. Lemma 3.5 gives that $\phi_n(x) = \mathbb{E}[\phi(X^\varepsilon((n-1)h,x))|x]$ is in $\mathcal{C}_P^3$. We rewrite the global error, given by $E_h^\varepsilon = |\mathbb{E}[\phi(X^\varepsilon(T,X_0)) - \phi(X_N^\varepsilon(X_0))]|$, with a telescopic sum,

$$\begin{aligned}
E_h^\varepsilon &\leqslant \sum_{n=1}^N \left|\mathbb{E}[\phi(X^\varepsilon(nh, X_{N-n}^\varepsilon(X_0))) - \phi(X^\varepsilon((n-1)h, X_{N-n+1}^\varepsilon(X_0)))]\right| \\
&\leqslant \sum_{n=1}^N \left|\mathbb{E}[\phi_n(X^\varepsilon(h, X_{N-n}^\varepsilon(X_0))) - \phi_n(X_1^\varepsilon(X_{N-n}^\varepsilon(X_0)))]\right| \\
&\leqslant \sum_{n=1}^N \mathbb{E}\left[\left|\mathbb{E}[\phi_n(X^\varepsilon(h,x)) - \phi_n(X_1^\varepsilon(x))|x = X_{N-n}^\varepsilon(X_0)]\right|\right] \\
&\leqslant \sum_{n=1}^N C(1 + \mathbb{E}[|X_{N-n}^\varepsilon(X_0)|^K])h^{3/2} \leqslant Ch^{1/2},
\end{aligned}$$

where we used the bounded moments property of Proposition 3.6 and $X_n^\varepsilon \in \mathcal{M}_h^\varepsilon$ (Lemma 3.1). $\square$

Proof of Theorem 2.9. We obtain straightforwardly from the expression of the integrator (2.2) that $|\zeta(X_n^\varepsilon)| \leqslant C_h\sqrt{\varepsilon}$. Using this estimate and the notation introduced in the proof of Lemma 3.8, we observe that

$$\left|Y_{(i)}^\varepsilon(X_n^\varepsilon) - Y_{(i)}^0(X_n^\varepsilon)\right| \leqslant C_h\sqrt{\varepsilon}, \quad \left|\zeta_{(i)}^\varepsilon(X_n^\varepsilon)\right| \leqslant C_h\sqrt{\varepsilon}, \quad i \in \{1/2, 1\},$$

where $Y_{(1/2)}^0(x) = \sigma\xi_n$ and $Y_{(1)}^0(x) = f(x)$. The Lagrange multiplier given by (3.16) therefore satisfies $\left|\lambda_{n+1}^\varepsilon - \widetilde{\lambda_{n+1}^\varepsilon}\right| \leqslant C_h\sqrt{\varepsilon}$, where

$$\widetilde{\lambda_{n+1}^\varepsilon} = -G_{X_{n+1}^\varepsilon - X_n^\varepsilon}^{-1}(X_n^\varepsilon) \int_0^1 g^T(X_n^\varepsilon + \tau(X_{n+1}^\varepsilon - X_n^\varepsilon))d\tau(\sigma\sqrt{h}\xi_n + hf(X_n^\varepsilon)).$$

Similarly to (3.16), the Lagrange multiplier λ_{n+1}^0 of the constrained Euler integrator (1.6) is given by

$$\lambda_{n+1}^0 = -G_{X_{n+1}^0 - X_n^0}^{-1}(X_n^0) \int_0^1 g^T(X_n^0 + \tau(X_{n+1}^0 - X_n^0))d\tau(\sigma\sqrt{h}\xi_n + hf(X_n^0)).$$

Using that $G_y^{-1}(x)$ is bounded, a straightforward calculation shows that $G_y^{-1}(x)$ is Lipschitz in $x, y \in \mathbb{R}^d$, that is, there exists a constant $L > 0$ such that

$$\left|G_{y_1}^{-1}(x_1) - G_{y_2}^{-1}(x_2)\right| \leqslant L(|x_1 - x_2| + |y_1 - y_2|), \quad x_1, x_2, y_1, y_2 \in \mathbb{R}^d.$$

19

Thus, we get the estimate

$$\left|\lambda_{n+1}^{\varepsilon} - \lambda_{n+1}^{0}\right| \leqslant C_h(\sqrt{\varepsilon} + \left|X_n^{\varepsilon} - X_n^0\right|) + C\sqrt{h}\left|X_{n+1}^{\varepsilon} - X_{n+1}^0\right|,$$

and, as $\lambda_{n+1}^{\varepsilon}$ and λ_{n+1}^{0} are bounded uniformly in ε, we have

$$\left|g(X_n^{\varepsilon})\lambda_{n+1}^{\varepsilon} - g(X_n^0)\lambda_{n+1}^{0}\right| \leqslant C_h(\sqrt{\varepsilon} + \left|X_n^{\varepsilon} - X_n^0\right|) + C\sqrt{h}\left|X_{n+1}^{\varepsilon} - X_{n+1}^0\right|.$$

From the definitions of X_{n+1}^{ε} in (2.2) and X_{n+1}^0 in (1.6), we deduce that

$$\left|X_{n+1}^{\varepsilon} - X_{n+1}^0\right| \leqslant C_h(\sqrt{\varepsilon} + \left|X_n^{\varepsilon} - X_n^0\right|) + C\sqrt{h}\left|X_{n+1}^{\varepsilon} - X_{n+1}^0\right|.$$

If $h \leqslant h_0$ is small enough, we obtain

$$\left|X_{n+1}^{\varepsilon} - X_{n+1}^0\right| \leqslant C_h(\sqrt{\varepsilon} + \left|X_n^{\varepsilon} - X_n^0\right|).$$

We deduce the estimate (2.6) by iterating this inequality for $n \leqslant N = T/h$. $\qquad\square$

Proof of Theorem 2.12. We take over the notations and the expression (3.16) of the Lagrange multiplier λ_1^{ε} that we used in the proof of Lemma 3.8. Without loss of generality, we concentrate on the first step of the algorithm with $X_0 = x$. Replacing λ_1^{ε} by the explicit formula (3.16) in (2.2) yields that X_1^{ε} is a fixed point of the following map:

$$F_h^{\varepsilon}(y) = x + \sqrt{h}Y_{(1/2)}^{\varepsilon}(x) + hY_{(1)}^{\varepsilon}(x) + g(x)G_{y-x}^{-1}(x)\Big[\sqrt{h}\zeta_{(1/2)}^{\varepsilon}(x) + h\zeta_{(1)}^{\varepsilon}(x)$$
$$- \int_0^1 g^T(x + \tau(y-x))d\tau(\sqrt{h}Y_{(1/2)}^{\varepsilon}(x) + hY_{(1)}^{\varepsilon}(x))\Big].$$

For y_1, $y_2 \in \mathbb{R}^d$, Assumption 2.1 gives

$$\left|F_h^{\varepsilon}(y_2) - F_h^{\varepsilon}(y_1)\right| \leqslant C\sqrt{h}\left|y_2 - y_1\right|,$$

where we used that $G_y^{-1}(x)$ is Lipschitz in x, $y \in \mathbb{R}^d$. We deduce that F_h^{ε} is a uniform contraction for $h \leqslant h_0$ small enough. $\qquad\square$

4 Numerical experiments

In this section, we perform numerical experiments to confirm the theoretical findings, on a torus in $\mathbb{R}^3$ and on the orthogonal group in high dimension and codimension, in the spirit of the experiments in [46, 47, 30].

4.1 Uniform approximation for the invariant measure on a torus

We consider the example in codimension one of a torus in $\mathbb{R}^3$. We apply the new method given by the discretization (2.4) for sampling the invariant measure of (1.4) for different steps h and parameters ε, and we compare it with the Euler integrators (1.5) in $\mathbb{R}^d$ and (1.6) on the manifold. We recall that for dynamics of the form (1.4), the weak convergence implies the convergence for the invariant measure (see [43]). The numerical experiments of this subsection hint that the uniform accuracy property also extends to the approximation of the invariant measure, but we leave the mathematical analysis for the invariant measure for future work. We consider the constraint $\zeta(x) = (x_1^2 + x_2^2 + x_3^2 + R^2 - r^2)^2 - 4R^2(x_1^2 + x_2^2)$, with $R = 3$ and $r = 1$,

and we choose the map $f(x) = -25(x_1 - R + r, x_2, x_3)$, with $\sigma = \sqrt{2}$, the test function $\phi(x) = |x|^2$, $M = 10^7$ trajectories, the final time $T = 10$, and the initial condition $X_0 = (R - r, 0, 0)$. Increasing the value of T does not modify the computed averages, which hints that we reached the equilibrium. The factor 25 in f confines the solution in a reasonably small neighborhood of the torus, which allows a faster convergence to equilibrium and to take fewer trajectories. We compute the Monte-Carlo estimator $\bar{J} = \frac{1}{M} \sum_{m=1}^{M} \phi(X_N^{(m)}) \simeq \mathbb{E}[\phi(X_N)]$, where $X_n^{(m)}$ is the m-th realization of the integrator at time $t_n = nh$, and N is an integer satisfying $Nh = T$. We compare this approximation with a reference value of $\int_{\mathbb{R}^d} \phi d\mu_\infty^\varepsilon$ computed with the uniformly accurate integrator, by using a timestep $h_{\text{ref}} = 2^{-12}$. In the case of the constrained Euler scheme (1.6), it amounts to comparing an approximation of $\int_{\mathcal{M}} \phi d\mu_\infty^0$ with the reference value of $\int_{\mathbb{R}^d} \phi d\mu_\infty^\varepsilon$. We observe in Figure 1 that the accuracy of the constrained integrator (1.6) for solving the unconstrained problem (1.4) deteriorates when ε grows larger, as μ_∞^ε deviates from μ_∞^0. The explicit Euler scheme (1.5) faces stability issues when $\varepsilon \to 0$. The accuracy of the new method for solving (1.4) does not deteriorate depending on ε, and it shares a behavior similar to the constrained Euler scheme (1.6) when $\varepsilon \to 0$, which is in agreement with Theorem 2.9. The right graph of Figure 1 shows that the behavior of the error in ε is the same for any fixed value of h. This is a numerical confirmation of the uniform accuracy property of the discretization (2.2) (in the spirit of the numerical experiments in [11]), as stated in Theorem 2.7. For any fixed ε, the right graph of Figure 1 also shows that the error decreases when $h \to 0$. A plot of the error against h for a fixed ε (not included for conciseness) shows a slope of order one (see Remark 2.8).
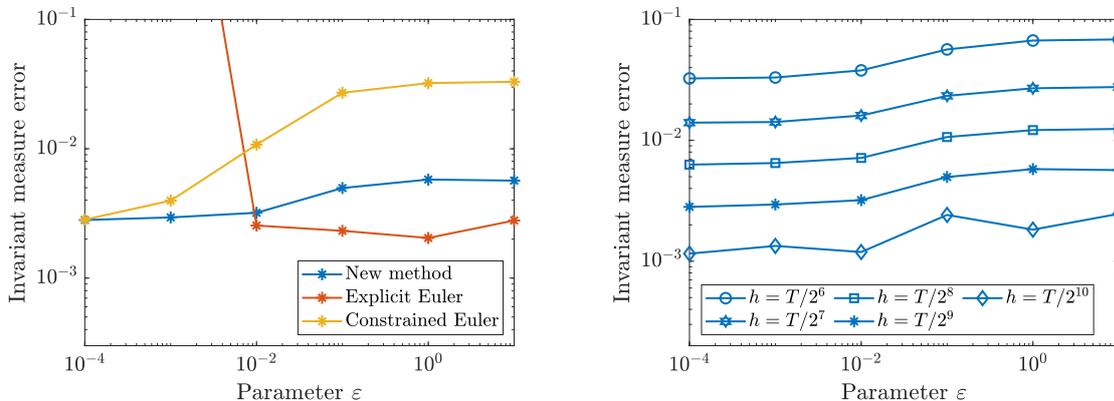

Figure 1: Error for sampling the invariant measure of penalized Langevin dynamics on a torus in $\mathbb{R}^3$ of the uniform discretization (2.2) and the Euler integrators (1.5) and (1.6) for different values of ε with $h = 2^{-9}T$ (left), and error curves versus ε of the uniformly accurate method for different timesteps $h = 2^{-i}T$ and $i = 6, \ldots, 10$ (right), with the final time $T = 10$, the maps $f(x) = -25(x_1 - R + r, x_2, x_3)$, $\phi(x) = |x|^2$, and $M = 10^7$ trajectories.

4.2   Weak approximation on the orthogonal group

We apply the uniformly accurate method on a compact Lie group (in the spirit of the numerical experiments in [46, 47]) to see how it performs in high dimension and codimension. We choose the orthogonal group $O(m) = \{M \in \mathbb{R}^{m \times m}, M^T M = I_m\}$, seen as a submanifold of $\mathbb{R}^{m^2}$ of codimension $q = m(m + 1)/2$. We compare the explicit Euler scheme (1.5), the constrained Euler scheme (1.6), and the new method on $\mathcal{M} = O(m)$ for $m = 2, \ldots, 5$ with

the parameters $\varepsilon = 0.005$, $T = 1$, and $h = 2^{-7}$. Note that, as h and ε share the same order of magnitude, the explicit Euler scheme (1.5) can face stability issues, and the solution does not lie on the manifold $\mathcal{M}$. Thus, we are in the regime where the convergence results for both Euler schemes (1.5)-(1.6) do not apply. We choose $f = -\nabla V$, where V is given by

$$V(x) = 50 \operatorname{Tr}((x - I_{m^2})^T (x - I_{m^2})) \tag{4.1}$$

with the parameters $\sigma = \sqrt{2}$, $X_0 = I_d$, $\phi(x) = \operatorname{Tr}(x)$, and $M = 10^6$ trajectories. The reference solution for $J(m) = \mathbb{E}[\phi(X^\varepsilon(T))]$ is computed with the uniformly accurate integrator with $h_{\mathrm{ref}} = 2^{-9}$. With the factor 50 in the potential (4.1), the trajectories stay close to I_{m^2}, and $J(m)$ is close to $\phi(I_{m^2}) = m$. This choice of factor permits one to explore a reasonably small area of $\mathrm{O}(m)$, in order to avoid zones close to $\mathcal{M}$ where the Gram matrix G has a bad condition number or is singular, and to reduce the number of trajectories needed. We observe numerically that replacing the factor 50 by 1 in (4.1) induces a severe timestep restriction. We present the results of the experiment in Table 1. We omit the results for the explicit Euler scheme (1.5) as the method is inaccurate in this regime (error of size 1). We observe that, in the regime where h and ε share the same order of magnitude, the uniformly accurate integrator performs significantly better than the Euler schemes (1.6) and (1.5) for solving the problem (1.4). In this regime, the Euler method (1.5) faces stability issues, and it is inappropriate to use the constrained Euler scheme (1.6) as the solution $X^\varepsilon(t)$ of (1.4) is not close to the solution $X^0(t)$ of (1.2). Moreover, the cost in time of the new method stays the same in average for any value of ε (results not included for conciseness). This confirms numerically the uniform cost of solving the fixed point problem (2.2), as stated in Theorem 2.12.

m	$\dim(\mathcal{M})$	q	$J(m)$	$\bar{J}_{\mathrm{UA}}$	error of $\bar{J}_{\mathrm{UA}}$	$\bar{J}_{\mathrm{EC}}$	error of $\bar{J}_{\mathrm{EC}}$
2	1	3	2.00934	2.00619	$3.1 \cdot 10^{-3}$	1.99165	$1.8 \cdot 10^{-2}$
3	3	6	3.01458	3.00821	$6.4 \cdot 10^{-3}$	2.97460	$4.0 \cdot 10^{-2}$
4	6	10	4.02050	4.00972	$1.1 \cdot 10^{-2}$	3.94846	$7.2 \cdot 10^{-2}$
5	10	15	5.02669	5.00842	$1.8 \cdot 10^{-2}$	4.91298	$1.1 \cdot 10^{-1}$

Table 1: Numerical approximation of $J(m) = \mathbb{E}[\phi(X^\varepsilon(T))]$ for $2 \leqslant m \leqslant 5$ with the estimator $\bar{J} = M^{-1} \sum_{k=1}^M \phi(X_N^{(k)})$, where (X_n) is given by the uniform discretization (2.2) for $\bar{J}_{\mathrm{UA}}$ and by the constrained Euler scheme (1.6) for $\bar{J}_{\mathrm{EC}}$ with their respective errors. The average is taken over $M = 10^6$ trajectories with the potential (4.1), $\phi(x) = \operatorname{Tr}(x)$, the final time $T = 1$, the stiff parameter $\varepsilon = 0.005$, and the timestep $h = 2^{-7}$.

5 Conclusion and future work

In this work, we presented a new method for the weak numerical integration of penalized Langevin dynamics evolving in the vicinity of manifolds of any dimension and codimension. On the contrary of the other existing discretizations, the accuracy of the proposed integrator is independent of the size of the stiff parameter ε. Moreover, its cost does not depend on ε, and it converges to the Euler scheme on the manifold when $\varepsilon \to 0$. Throughout the analysis, we gave an expansion in time of the solution to the penalized Langevin dynamics that is uniform in ε, as well as new tools for the study of stochastic projection methods for solving stiff SDEs.

Multiple questions arise from the work presented in this paper, with many of great interest for physical applications. First, it would be interesting to get convergence results with weaker assumptions or to develop uniformly accurate integrators for different penalized dynamics with the same limit when $\varepsilon \to 0$ such as the original penalized dynamics (1.3) (see [16]). One could build integrators for penalized dynamics of the form

$$dX^\varepsilon = f(X^\varepsilon)dt + \sigma dW + \frac{\sigma^2}{4}\nabla \ln(\det(G)) - \frac{1}{\varepsilon}(gG^{-1}\zeta_1)(X^\varepsilon)dt - \frac{1}{\nu}(gG^{-1}\zeta_2)(X^\varepsilon)dt,$$

where ε and ν do not share the same order of magnitude, or for constrained dynamics with a penalized term. One could also build a uniformly accurate numerical scheme with high order in the weak context, or just in the context of the invariant measure (in the spirit of the works [6, 31, 2, 3, 32, 28, 30] where numerical schemes of high order for the invariant measure and weak order one were introduced). Postprocessors [45] proved to be an efficient tool for reaching high order for the invariant measure without increasing the cost of the method and could be used in this context. Moreover, the order conditions presented in [28] for Runge-Kutta methods for solving Langevin dynamics in $\mathbb{R}^d$ both in the weak sense and for the invariant measure do not match with the order conditions for solving Langevin dynamics constrained on the manifold $\mathcal{M}$, as presented in [30]. It would be interesting to create a unified class of high order Runge-Kutta methods with the same order conditions in $\mathbb{R}^d$, on the manifold $\mathcal{M}$ and in the vicinity of $\mathcal{M}$. The discretizations presented in this paper could also be combined with Metropolis-Hastings rejection procedures [39, 24], in the spirit of the works [20, 8, 35, 46, 36], in order to get an exact approximation for the invariant measure with a rejection rate that does not deteriorate in the regime $\varepsilon \to 0$.

References

[1] A. Abdulle, D. Cohen, G. Vilmart, and K. C. Zygalakis. High weak order methods for stochastic differential equations based on modified equations. *SIAM J. Sci. Comput.*, 34(3):A1800–A1823, 2012.

[2] A. Abdulle, G. Vilmart, and K. C. Zygalakis. High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Numer. Anal.*, 52(4):1600–1622, 2014.

[3] A. Abdulle, G. Vilmart, and K. C. Zygalakis. Long time accuracy of Lie-Trotter splitting methods for Langevin dynamics. *SIAM J. Numer. Anal.*, 53(1):1–16, 2015.

[4] I. Almuslimani, P. Chartier, M. Lemou, and F. Méhats. Uniformly accurate schemes for oscillatory stochastic differential equations. *Submitted*, 2021.

[5] H. C. Andersen. Rattle: A "velocity" version of the Shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52(1):24–34, 1983.

[6] N. Bou-Rabee and H. Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Numer. Anal.*, 48(1):278–297, 2010.

[7] C.-E. Bréhier and S. Rakotonirina-Ricquebourg. On asymptotic preserving schemes for a class of stochastic differential equations in averaging and diffusion approximation regimes. *Multiscale Model. Simul.*, 20(1):118–163, 2022.

[8] M. Brubaker, M. Salzmann, and R. Urtasun. A family of MCMC methods on implicitly defined manifolds. In *Artificial intelligence and statistics*, pages 161–172, 2012.

[9] M. Calvo, L. O. Jay, J. I. Montijano, and L. Rández. Approximate compositions of a near identity map by multi-revolution Runge-Kutta methods. *Numer. Math.*, 97(4):635–666, 2004.

[10] M. Calvo, J. I. Montijano, and L. Rández. On explicit multi-revolution Runge-Kutta schemes. *Adv. Comput. Math.*, 26(1-3):105–120, 2007.

[11] P. Chartier, N. Crouseilles, M. Lemou, and F. Méhats. Uniformly accurate numerical schemes for highly oscillatory Klein-Gordon and nonlinear Schrödinger equations. *Numer. Math.*, 129(2):211–250, 2015.

[12] P. Chartier, M. Lemou, F. Méhats, and G. Vilmart. A new class of uniformly accurate numerical schemes for highly oscillatory evolution equations. *Found. Comput. Math.*, 20(1):1–33, 2020.

[13] P. Chartier, M. Lemou, and L. Trémant. Uniformly accurate numerical schemes for a class of dissipative systems. *Submitted*, 2020.

[14] P. Chartier, J. Makazaga, A. Murua, and G. Vilmart. Multi-revolution composition methods for highly oscillatory differential equations. *Numer. Math.*, 128(1):167–192, 2014.

[15] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem*, 6(9):1809–1814, 2005.

[16] G. Ciccotti, T. Lelievre, and E. Vanden-Eijnden. Projection of diffusions on submanifolds: application to mean force computation. *Comm. Pure Appl. Math.*, 61(3):371–408, 2008.

[17] D. Cohen and M. Sigg. Convergence analysis of trigonometric methods for stiff second-order stochastic differential equations. *Numer. Math.*, 121(1):1–29, 2012.

[18] A. Debussche and E. Faou. Weak backward error analysis for SDEs. *SIAM J. Numer. Anal.*, 50(3):1735–1752, 2012.

[19] I. I. Gikhman and A. V. Skorokhod. Stochastic differential equations. In *The Theory of Stochastic Processes III*, pages 113–219. Springer, 2007.

[20] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[21] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. Structure-preserving algorithms for ordinary differential equations.

[22] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010. Stiff and differential-algebraic problems, Second revised edition, paperback.

[23] R. Z. Hasminskii. *Stochastic stability of differential equations*, volume 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn—Germantown, Md., 1980. Translated from the Russian by D. Louvish.

[24] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[25] S. Jin. Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. *Riv. Math. Univ. Parma (N.S.)*, 3(2):177–216, 2012.

[26] M. Kopec. Weak backward error analysis for Langevin process. *BIT Numer. Math.*, 55(4):1057–1103, 2015.

[27] M. Kopec. Weak backward error analysis for overdamped Langevin processes. *IMA J. Numer. Anal.*, 35(2):583–614, 2015.

[28] A. Laurent and G. Vilmart. Exotic aromatic B-series for the study of long time integrators for a class of ergodic SDEs. *Math. Comp.*, 89(321):169–202, 2020.

[29] A. Laurent and G. Vilmart. Multirevolution integrators for differential equations with fast stochastic oscillations. *SIAM J. Sci. Comput.*, 42(1):A115–A139, 2020.

[30] A. Laurent and G. Vilmart. Order conditions for sampling the invariant measure of ergodic stochastic differential equations on manifolds. *Found. Comput. Math.*, 22(3):649–695, 2022.

[31] B. Leimkuhler and C. Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. Express. AMRX*, 2013(1):34–56, 2013.

[32] B. Leimkuhler, C. Matthews, and G. Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.*, 36(1):13–79, 2016.

[33] T. Lelièvre, C. Le Bris, and E. Vanden-Eijnden. Analyse de certains schémas de discrétisation pour des équations différentielles stochastiques contraintes. *C. R. Math. Acad. Sci. Paris*, 346(7-8):471–476, 2008.

[34] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations*. Imperial College Press, London, 2010. A mathematical perspective.

[35] T. Lelièvre, M. Rousset, and G. Stoltz. Langevin dynamics with constraints and computation of free energy differences. *Math. Comp.*, 81(280):2071–2125, 2012.

[36] T. Lelièvre, M. Rousset, and G. Stoltz. Hybrid Monte Carlo methods for sampling probability measures on submanifolds. *Numer. Math.*, 143(2):379–421, 2019.

[37] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[38] B. Melendo and M. Palacios. A new approach to the construction of multirevolution methods and their implementation. *Appl. Numer. Math.*, 23(2):259–274, 1997.

[39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[40] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics.* Scientific Computation. Springer-Verlag, Berlin, 2004.

[41] P. Plecháč and M. Rousset. Implicit mass-matrix penalization of Hamiltonian dynamics with application to exact sampling of stiff systems. *Multiscale Model. Simul.*, 8(2):498–539, 2009/10.

[42] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.

[43] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

[44] G. Vilmart. Weak second order multirevolution composition methods for highly oscillatory stochastic differential equations with additive or multiplicative noise. *SIAM J. Sci. Comput.*, 36(4):A1770–A1796, 2014.

[45] G. Vilmart. Postprocessed integrators for the high order integration of ergodic SDEs. *SIAM J. Sci. Comput.*, 37(1):A201–A220, 2015.

[46] E. Zappa, M. Holmes-Cerfon, and J. Goodman. Monte Carlo on manifolds: sampling densities and integrating functions. *Comm. Pure Appl. Math.*, 71(12):2609–2647, 2018.

[47] W. Zhang. Ergodic SDEs on submanifolds and related numerical sampling schemes. *ESAIM Math. Model. Numer. Anal.*, 2019.

[48] K. C. Zygalakis. On the existence and the applications of modified equations for stochastic differential equations. *SIAM J. Sci. Comput.*, 33(1):102–130, 2011.

Appendices

A   Proof of Theorem 2.3

In this section, we prove the convergence of the penalized Langevin dynamics (1.4) to the constrained dynamics (1.2) when $\varepsilon \to 0$, as stated in Theorem 2.3. The proof uses techniques and arguments similar to those in [16, Appx. C]. However, since we rescaled the stiff term in (1.4), there is no need for a change of time to prove the convergence to the constrained dynamics.

Proof of Theorem 2.3. In the orthogonal coordinates system given by Assumption 2.1, equation (1.4) becomes

$$d\varphi(X^\varepsilon(t)) = (\varphi'f)(X^\varepsilon(t))dt + \sigma\varphi'(X^\varepsilon(t))dW(t)$$

$$+ \frac{\sigma^2}{4}(\varphi' \nabla \ln(\det(G)))(X^\varepsilon(t))dt + \frac{\sigma^2}{2}\Delta\varphi(X^\varepsilon(t))dt,$$

$$d\zeta(X^\varepsilon(t)) = (g^T f)(X^\varepsilon(t))dt + \sigma g^T(X^\varepsilon(t))dW(t) + \frac{\sigma^2}{4}(g^T \nabla \ln(\det(G)))(X^\varepsilon(t))dt$$

$$+ \frac{\sigma^2}{2}\operatorname{div}(g)(X^\varepsilon(t))dt - \frac{1}{\varepsilon}\zeta(X^\varepsilon(t))dt,$$

where we used that $\varphi' g = 0$ and that $\zeta' = g^T$. Therefore, $\zeta(X^\varepsilon(t))$ satisfies

$$\zeta(X^\varepsilon(t)) = \sigma \int_0^t e^{(s-t)/\varepsilon} g^T(X^\varepsilon(s))dW$$

$$+ \int_0^t e^{(s-t)/\varepsilon}\Big[g^T f + \frac{\sigma^2}{4}g^T \nabla \ln(\det(G)) + \frac{\sigma^2}{2}\operatorname{div}(g)\Big](X^\varepsilon(s))dt,$$

and a bound on $\zeta(X^\varepsilon(t))$ is given by

$$\mathbb{E}[|\zeta(X^\varepsilon(t))|^2] \leqslant C \int_0^t e^{2(s-t)/\varepsilon}dt + C\Big(\int_0^t e^{(s-t)/\varepsilon}dt\Big)^2 \leqslant C\varepsilon. \qquad \text{(A.1)}$$

On the other hand, we rewrite (1.2) in the orthogonal coordinates as

$$d\varphi(X^0(t)) = (\varphi' f)(X^0(t))dt + \sigma\varphi'(X^0(t))dW$$

$$+ \frac{\sigma^2}{4}(\varphi' \nabla \ln(\det(G)))(X^0(t))dt + \frac{\sigma^2}{2}\Delta\varphi(X^0(t))dt,$$

$$d\zeta(X^0(t)) = 0,$$

where we used that

$$\frac{1}{2}\varphi' \nabla \ln(\det(G)) + \Delta\varphi = \sum_{i=1}^d \varphi'(\Pi'_{\mathcal{M}}(\Pi_{\mathcal{M}}e_i)e_i) + \sum_{i=1}^d \varphi''(\Pi_{\mathcal{M}}e_i, \Pi_{\mathcal{M}}e_i).$$

With Assumption 2.1, we obtain

$$\mathbb{E}[|\varphi(X^\varepsilon(t)) - \varphi(X^0(t))|^2] \leqslant C \int_0^t \mathbb{E}[|X^\varepsilon(s) - X^0(s)|^2]ds,$$

and from (A.1) we deduce that

$$\mathbb{E}[|\zeta(X^\varepsilon(t)) - \zeta(X^0(t))|^2] \leqslant C\varepsilon.$$

Thus, $\psi(X^\varepsilon(t))$ satisfies

$$\mathbb{E}[|\psi(X^\varepsilon(t)) - \psi(X^0(t))|^2] \leqslant C \int_0^t \mathbb{E}[|X^\varepsilon(s) - X^0(s)|^2]ds + C\varepsilon,$$

and, as ψ^{-1} is Lipschitz, we have

$$\mathbb{E}[|X^\varepsilon(t) - X^0(t)|^2] \leqslant C \int_0^t \mathbb{E}[|X^\varepsilon(s) - X^0(s)|^2]ds + C\varepsilon,$$

which, with the use of the Gronwall lemma, gives the desired estimate. $\qquad \square$