High order numerical simulation of the underdamped Langevin diffusion

James Foster

Supervised by Terry Lyons and Harald Oberhauser

University of Oxford

2019

Introduction

The underdamped Langevin diffusion is given by the following SDE:

$$dQ_t = P_t \, dt, \tag{1}$$

$$dP_t = -\nabla U(Q_t)\, dt - \nu P_t \, dt + \sqrt{\frac{2\nu}{\beta}}\, dW_t, \tag{2}$$

where $Q, P \in \mathbb{R}^d$ will represent the position and velocity of a particle moving in a potential $U : \mathbb{R}^d \to \mathbb{R}$ under the influence of a frictional force (with coefficient $\nu > 0$) along with a stochastic force given by a d-dimensional Brownian motion W (and inverse temperature β).

The underdamped Langevin equation can be viewed as an extension of Newton's second law and is a key model in statistical mechanics. ([1] is a detailed textbook that surveys these scientific applications)

Introduction

Recently this equation has been applied to computational statistics as, under mild conditions on the potential (see [2] for more details), the process admits a stationary measure with the following density:

$$\varphi(q,p) \propto \exp\left(-\beta\left(U(q) + \frac{1}{2}\|p\|^2\right)\right).$$

Hence by setting U equal to the log density of a target distribution, it is possible to generate samples by simulating a Langevin diffusion.

In practice, (1) can not be solved exactly and must be approximated.

Introduction

Our strategy for discretizing the underdamped Langevin equation is

1. Replace the Brownian motion W by a piecewise linear path $\widehat{W}$.

2. Along each piece of $\widehat{W}$ we can approximate (1) using the ODE:

$$d\widehat{Q}_t = \widehat{P}_t\,dt, \tag{3}$$

$$d\widehat{P}_t = f\big(\widehat{Q}_t\big)\,dt - \nu\widehat{P}_t\,dt + \sigma\,d\widehat{W}_t, \tag{4}$$

   where $f := -\nabla U$ and $\sigma := \sqrt{\frac{2\nu}{\beta}}$.

3. In each step, discretize (3) and (4) with a suitable ODE solver.

Piecewise linear discretization of Brownian motion

To construct the path $\widehat{W}$, we shall use the below theorem from [3].

Theorem (Brownian motion as a cubic with independent noise)

Consider a standard Brownian motion W over the unit interval $[0,1]$. Let $\widetilde{W}$ be the (unique) cubic polynomial with a root at 0 such that

$$\int_0^1 u^k \, d\widetilde{W}_u = \int_0^1 u^k \, dW_u, \quad \text{for } \; k = 0, 1, 2.$$

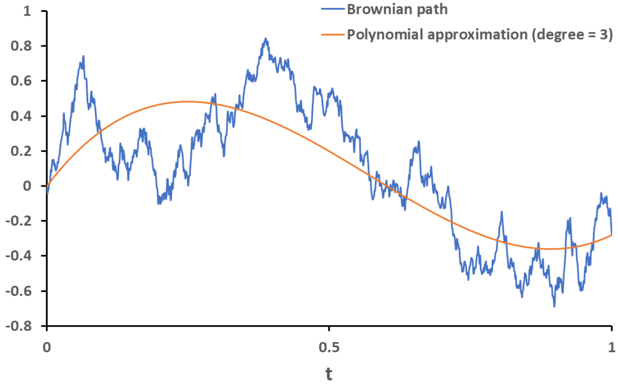
Then

$$W = \widetilde{W} + Z,$$

where Z is a centred Gaussian process independent of the cubic $\widetilde{W}$.

Piecewise linear discretization of Brownian motion

Therefore, the cubic polynomial $\widetilde{W}$ is an unbiased estimator for W.



The piecewise linear path $\widehat{W}$ will be designed to resemble this cubic.

Piecewise linear discretization of Brownian motion

Theorem (Pathwise approximation of the cubic polynomial $\widetilde{W}$)

There exists a piecewise linear path $\widehat{W}$ on $[0,1]$ and $a > 0$ such that:

1. $\widehat{W}_0 = 0$.
2. $\widehat{W}$ *has three pieces which connect at* $\frac{1}{2}\left(1 \pm a\right)$.
3. $\displaystyle\int_0^1 u^k \, d\widehat{W}_u = \int_0^1 u^k \, d\widetilde{W}_u$, *for* $k = 0, 1, 2, 3$.

Proof.

Note that for any fixed $a \in \left(0, 1\right)$, the first three integral conditions and $\widehat{W}_0$ are enough to uniquely determine the piecewise linear path.

Hence, the idea of the proof is to derive a necessary condition for a.

Piecewise linear discretization of Brownian motion

Proof. (continued)

Using the orthogonality of the Legendre polynomials, it follows that

$$\begin{aligned}\int_0^1 \left(20u^3 - 30u^2 + 12u - 1\right) d\widehat{W}_u &= \int_0^1 \left(20u^3 - 30u^2 + 12u - 1\right) d\widetilde{W}_u \\ &= \int_0^1 \left(20u^3 - 30u^2 + 12u - 1\right)\left(\widetilde{W}_u\right)' du \\ &= 0.\end{aligned}$$

On the other hand, the above integral can be explicitly computed as

$$\begin{aligned}\int_0^1 \left(20u^3 - 30u^2 + 12u - 1\right) d\widehat{W}_u \\ = \frac{1}{8}\left(5a^3 + 5a^2 - a - 1\right)\left(\widehat{W}_1 - \widehat{W}_{\frac{1}{2}(1+a)} - \widehat{W}_{\frac{1}{2}(1-a)}\right).\end{aligned}$$

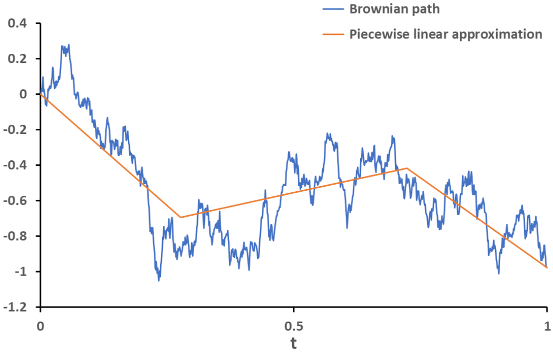
Piecewise linear discretization of Brownian motion

Proof. (continued)

So in order for $\widehat{W}$ to exist, the constant a must satisfy the equation

$$5a^3 + 5a^2 - a - 1 = 0.$$

Therefore by setting $a = \frac{\sqrt{5}}{5}$, we can construct the desired path. $\square$



Piecewise linear discretization of Brownian motion

In practice, $\widehat{W}$ can be generated on $[s,t]$ using the random variables

$$W_{s,t} := W_t - W_s,$$
$$H_{s,t} := \frac{1}{h}\int_s^t W_{s,u} - \frac{u-s}{h}W_{s,t}\,du,$$
$$K_{s,t} := \frac{1}{h^2}\int_s^t \left(W_{s,u} - \frac{u-s}{h}W_{s,t}\right)\left(\frac{1}{2}h - (u-s)\right)du,$$

where $h = t - s$.

It is straightforward to prove that $\left(W_{s,t}, H_{s,t}, K_{s,t}\right)$ are independent and uniquely determine the cubic polynomial $\widetilde{W}$ (see [3] for details).

Furthermore $W_{s,t} \sim N\big(0,h\big)$, $H_{s,t} \sim N\big(0,\frac{1}{12}h\big)$ and $K_{s,t} \sim N\big(0,\frac{1}{720}h\big)$.

Piecewise linear discretization of Brownian motion

Theorem (Pathwise approximation of the Brownian motion W)

Let $\widehat{W}$ be the piecewise linear path on $[s,t]$ with three pieces which connect the points (s, W_s), $(s + Ah, B)$, $(t - Ah, C)$, (t, W_t) given by

$$A = \frac{1}{2} - \frac{1}{2\sqrt{5}},$$

$$\frac{1}{2}\big(B + C\big) = W_s + \frac{1}{2}W_{s,t} + \frac{1}{1-A}H_{s,t},$$

$$\frac{1}{2}\big(B - C\big) = \frac{6}{1-A}K_{s,t} - \frac{1}{2}\big(1-2A\big)W_{s,t}.$$

Then $\widehat{W}$ matches the increment $W_{s,t}$, the time areas $(H_{s,t}, K_{s,t})$ and

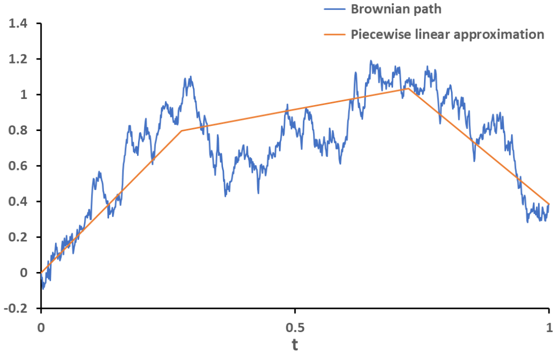
$$\int_s^t (u-s)^2\,\widehat{W}_{s,u}\,du = \mathbb{E}\left[\int_s^t (u-s)^2\,W_{s,u}\,du \,\middle|\, W_{s,t}, H_{s,t}, K_{s,t}\right].$$

Piecewise linear discretization of Brownian motion

Proof.

The result is a direct consequence of the previous two theorems. □

When approximating the underdamped Langevin diffusion on $[0, T]$, we shall generate $\widehat{W}$ on the uniform grid $\triangle_N$ with mesh size $h = \frac{T}{N}$.



Approximation of underdamped Langevin diffusion by ODEs

Recall that the underdamped Langevin equation is the SDE given by

$$dQ_t = P_t\,dt, \tag{5}$$

$$dP_t = f(Q_t)\,dt - \nu P_t\,dt + \sigma\,dW_t. \tag{6}$$

Provided that f is twice differentiable, we have the Taylor expansion

$$\begin{aligned}\begin{pmatrix} Q_t \\ P_t \end{pmatrix} = \begin{pmatrix} Q_s \\ P_s \end{pmatrix} &+ (\cdots)\,h + (\cdots)\,W_{s,t} + (\cdots)\int_s^t W_{s,u}\,du + (\cdots)\,h^2 + (\cdots)\,h^3 \\ &+ (\cdots)\int_s^t\int_s^u W_{s,v}\,dv\,du + (\cdots)\int_s^t\int_s^u\int_s^v W_{s,r}\,dr\,dv\,du \\ &+ (\cdots)\int_s^t\int_s^u\int_s^v (r-s)\,dW_r\,dv\,du + O(h^4),\end{aligned}$$

where $(\cdots)$ are terms involving the vector fields and their derivatives.

Approximation of underdamped Langevin diffusion by ODEs

Similarly, $(\widehat{Q}, \widehat{P})$ will also have this expansion but with $\widehat{W}$ integrals.

By the previous theorem, these integrals are approximated optimally using $\widehat{W}$ (in an $L^2(\mathbb{P})$ sense) and hence gives the local error estimate

Theorem

Provided the vector field f and its first two derivatives are bounded and Lipschitz continuous, there exist constants $C_1, C_2 > 0$ such that

$$\mathbb{E}\left[\left\|\begin{pmatrix} Q_t \\ P_t \end{pmatrix} - \begin{pmatrix} \widehat{Q}_t \\ \widehat{P}_t \end{pmatrix}\right\|^2\right] \leq \left(1 + C_1 h\right)\mathbb{E}\left[\left\|\begin{pmatrix} Q_s \\ P_s \end{pmatrix} - \begin{pmatrix} \widehat{Q}_s \\ \widehat{P}_s \end{pmatrix}\right\|^2\right] + C_2 h^7, \quad (7)$$

for sufficiently small h.

Approximation of underdamped Langevin diffusion by ODEs

Sketch Proof.

We derive (7) from the Taylor expansions of $\big(Q_t, P_t\big)$ and $\big(\widehat{Q}_t, \widehat{P}_t\big)$.

The first term in the estimate follows using the Lipschitz continuity.

Therefore it is helpful to consider the case when $\big(Q_s, P_s\big) = \big(\widehat{Q}_s, \widehat{P}_s\big)$. As the $O\big(h^{\frac{7}{2}}\big)$ integrals in both expansions have mean zero, we have

$$\mathbb{E}\left[\begin{pmatrix} Q_t \\ P_t \end{pmatrix} - \begin{pmatrix} \widehat{Q}_t \\ \widehat{P}_t \end{pmatrix}\right] \sim O(h^4). \tag{8}$$

It's important that the above quantity is strictly smaller than $O\big(h^{\frac{7}{2}}\big)$ so that large (but unbiased) terms disappear in the full expansion of

$$\mathbb{E}\left[\left\|\begin{pmatrix} Q_t \\ P_t \end{pmatrix} - \begin{pmatrix} \widehat{Q}_t \\ \widehat{P}_t \end{pmatrix}\right\|^2\right].$$

Sketch Proof. (continued)

The explicit calculation required to derive (7) is lengthy but follows the typical argument for $L^2(\mathbb{P})$ error estimation (see [4] for details).

The key idea is that the leading error terms are $O\big(h^{\frac{7}{2}}\big)$ but unbiased. So by squaring and taking an expectation, the terms that cannot be estimated using Lipschitz continuity will have a size of $O\big(h^7\big)$. $\square$

The above theorem then leads to the following global error estimate

Theorem (Strong convergence rate of the ODE approximation)

Suppose that $\big(\widehat{Q},\widehat{P}\big)$ *was obtained on* $[0,T]$ *using a uniform grid* $\triangle_N$ *with mesh size* $h=\frac{T}{N}$ *and an initial value* $\big(Q_0,P_0\big)=\big(\widehat{Q}_0,\widehat{P}_0\big)$*. Then*

$$\left\|\begin{pmatrix}Q_T\\ P_T\end{pmatrix}-\begin{pmatrix}\widehat{Q}_T\\ \widehat{P}_T\end{pmatrix}\right\|_{L^2(\mathbb{P})}\sim O\big(h^3\big).$$

Approximation of underdamped Langevin diffusion by ODEs

Similarly, by comparing the two Taylor expansions we can show that

Theorem (Weak convergence rate of the ODE approximation)

Suppose that $(\widehat{Q}, \widehat{P})$ *was obtained on* $[0, T]$ *using a uniform grid* $\triangle_N$ *with a mesh size of* $h = \frac{T}{N}$ *and the initial value* $(Q_0, P_0) = (\widehat{Q}_0, \widehat{P}_0)$*. Then for any polynomial* p*, there exists a constant* $C_p > 0$ *such that*

$$\left|\mathbb{E}\left[p\begin{pmatrix} Q_T \\ P_T \end{pmatrix}\right] - \mathbb{E}\left[p\begin{pmatrix} \widehat{Q}_T \\ \widehat{P}_T \end{pmatrix}\right]\right| \leq C_p h^3,$$

for all sufficiently small step sizes h*.*

Remark

The ODE driven by $\widehat{W}$ gives a third order numerical method for (1).

Discretization of the underdamped Langevin ODE

Since the ODE driven by $\widehat{W}$ is a third order approximation, it should be discretized using a numerical method that is at least third order.

Due to the structure of (3, 4), we can apply Runge-Kutta-Nyström (RKN) methods, which are explicitly designed for ODEs of the form

$$
\begin{aligned}
y'' &= F(t, y), \\
y(0) = y_0, &\quad y'(0) = y_0',
\end{aligned}
\tag{9}
$$

In addition, we wish for the RKN method to be symplectic since the underdamped Langevin equation is a stochastic Hamiltonian system.

Fortunately, the RKN method proposed by [5] has these properties.

Discretization of the underdamped Langevin ODE

To propagate the numerical solution over $[s,t]$, we use the variables

$$\begin{pmatrix} y_u \\ x_u \end{pmatrix} := \begin{pmatrix} e^{\frac{1}{2}\nu(u-s)}\widehat{Q}_u \\ e^{\nu(u-s)}\widehat{P}_u \end{pmatrix}, \qquad \forall u \in [s,t],$$

to represent the system (3, 4) using the following second order ODE

$$\frac{d^2 y}{du^2} = \frac{1}{4}\nu^2 y + e^{\frac{1}{2}\nu(u-s)}\left(f\big(e^{-\frac{1}{2}\nu(u-s)}y\big) + \sigma\,\frac{d\widehat{W}}{du}\right), \tag{10}$$

$$y_s = \widehat{Q}_s, \qquad y_s' = \widehat{P}_s + \frac{1}{2}\nu\,\widehat{Q}_s.$$

We can apply RKN methods on the intervals where $\frac{d\widehat{W}}{du}$ is constant.

Discretization of the underdamped Langevin ODE

For the numerical example, we will use the below numerical method.

Definition (A third order three-stage symplectic RKN method)

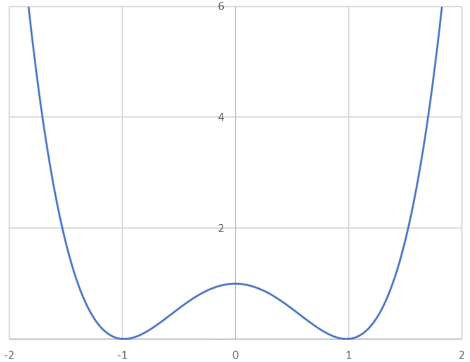
This method is presented in [5] and defined by the Butcher tableau:

$$
\begin{array}{c|ccc}
0 & & & \\
0.630847693 & 0.164217030 & & \\
0.536704894 & 0.139710559 & -0.103005664 & \\
\hline
 & 0.260311692 & 0.4039053382 & -0.164217030 \\
 & 0.260311692 & 1.094142798 & -0.354454490
\end{array}
$$

Numerical example: A particle in a double-well potential

For the numerical experiment, we shall consider a scalar double-well potential with the same parameters used in the first example of [6]:

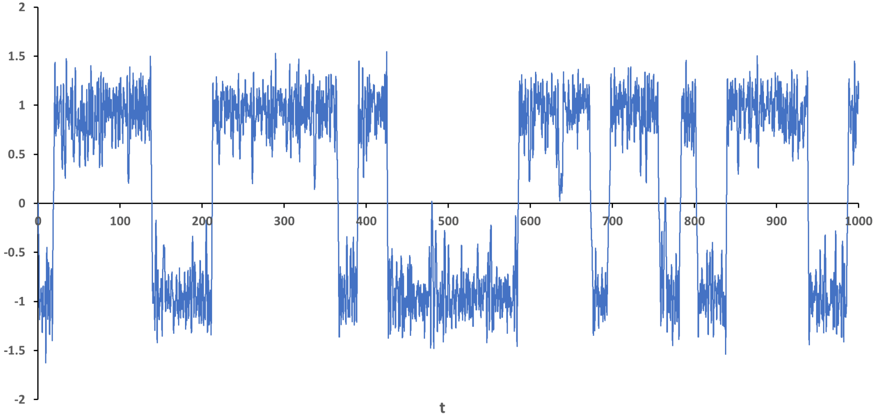
$$U(q) = \left(q^2 - 1\right)^2,$$



$$\nu = 1, \quad \beta = 3, \quad \left(Q_0, P_0\right) = 0.$$

Numerical example: A particle in a double-well potential

A sample path of the Langevin diffusion obtained by discretizing (3)



Numerical example: A particle in a double-well potential

Let $\widehat{Q}_T$ be the approximate position at time T computed with $h = \frac{T}{N}$.

We examine the strong and weak convergence using the estimators:

$$S_N := \sqrt{\mathbb{E}\Big[\big(\widehat{Q}_T - Q_T^{\text{fine}}\big)^2\Big]}, \tag{11}$$

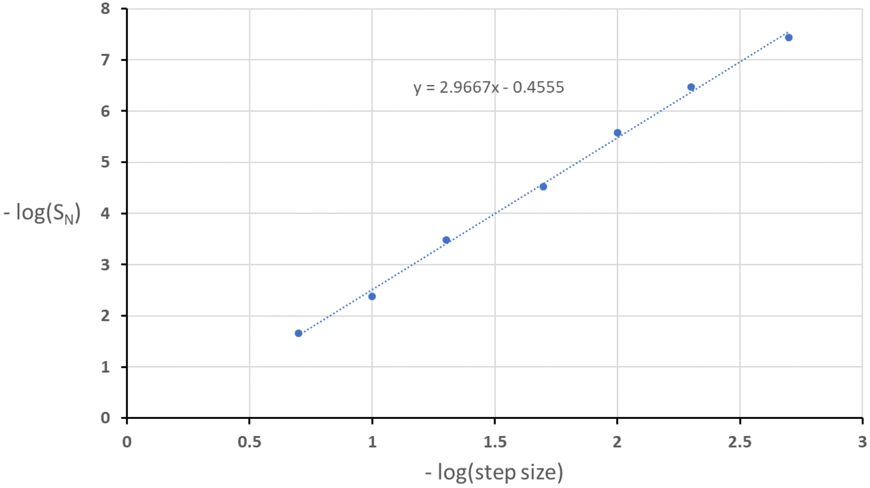
$$E_N := \Big|\mathbb{E}\big[\widehat{Q}_T^2\big] - \mathbb{E}\big[\big(Q_T^{\text{fine}}\big)^2\big]\Big|, \tag{12}$$

where the expectations are approximated by Monte-Carlo simulation and Q_T^{fine} denotes the numerical solution of (1) obtained at time T using the proposed method with a “fine” step size of $\min\left(\frac{h}{25}, \frac{T}{50000}\right)$.

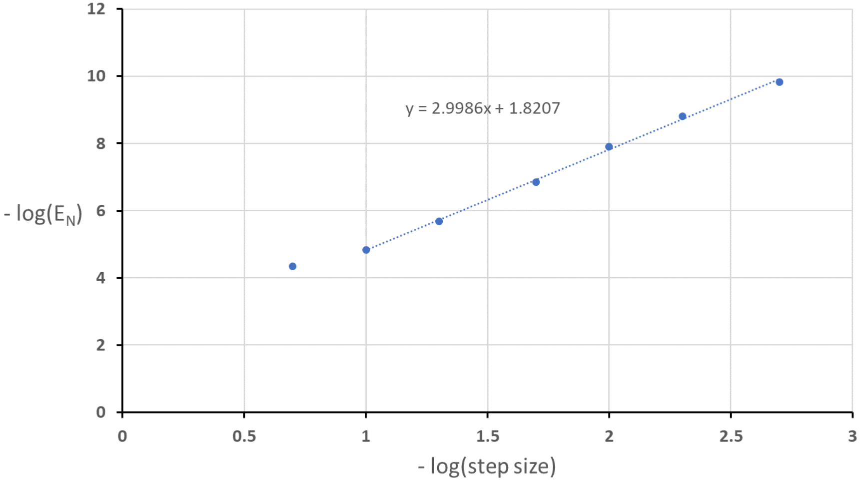
We will compute both $\widehat{Q}_T$ and Q_T^{fine} using the same Brownian paths.

For the numerical experiment, we shall use a time horizon of $T = 10$.

Numerical example: A particle in a double-well potential



Numerical example: A particle in a double-well potential



Future work

- How well does the proposed numerical method approximate the stationary distribution of the underdamped Langevin diffusion?
- Is this method practical for problems within Bayesian statistics?
- What are best ODE solvers for discretizing the system (3), (4)?
- Can this methodology be improved by using variable step sizes?

Thank you
for your attention!

References I

[1] W. T. Coffey, Y. P. Kalmykov and J. T. Waldron, *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, World Scientifc, 2012.

[2] X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan, *Underdamped Langevin MCMC: A non-asymptotic analysis*, Proceedings of the 31st Conference On Learning Theory, Volume 75 of Proceedings of Machine Learning Research, 2018.

[3] J. Foster, T. Lyons and H. Oberhauser, *An optimal polynomial approximation of Brownian motion*, arXiv:1904.06998, 2019.

References II

[4] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.

[5] M. A. Demba, N. Senu and F. Ismail, *A symplectic explicit trigonometrically-fitted Runge-Kutta-Nyström method for the numerical solution of periodic problems*, International Journal of Applied Engineering Research, Volume 11, 2016.

[6] N. Vercauteren, *Numerical investigation of solutions of Langevin equations*, Master's thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, 2005.