Weak Convergence and Long-Time Accuracy in Brownian Dynamics

Axel Bergengren*, Zhiyuan Liu†

KTH Royal Institute of Technology, axel@bergengren.com* Southern University of Science and Technology, 12311406@mail.sustech.edu.cn†

Introduction and Historical Background

This paper studies numerical schemes for stochastic gradient systems: SDE in d dimensions with the form

$$dX = a(X)dt + \sigma dw, \quad X(0) = X_0,$$

where $a := -\nabla V$, $V(x)$ is a potential energy function, and $\sigma > 0$ is a constant characterizing the strength of noise.
Previous work has shown the non-Markovian scheme

$$X_{k+1} = X_k + h\,a(X_k) + \sigma\frac{\sqrt{h}}{2}(\xi_k + \xi_{k+1}),$$

which, although only first-order weakly accurate on finite time intervals, produces a second-order accurate distribution in the limit as $t \to \infty$. This paper studies the transition from $O(h)$ to $O(h^2)$ accuracy, and proves the $O(h)$ term declines exponentially with time.

Preliminaries and Assumptions

Significant Assumptions

1. Potential Energy Condition: $V(x) \in C^7(\mathbb{R}^d)$, there exists constant $K > 0$
$$|a(x)| \le K(1+|x|) \quad |D^{\mathbf{i}}V(x)| \le K,$$
for all $\mathbf{i} = (i_1, i_2, \cdots, i_d), 2 \le |\mathbf{i}| \le 7$.
2. Observation Function Condition: Observation function $\varphi(x) \in C^6(\mathbb{R}^d)$, there exists one polynomial P such that
$$|D^{\mathbf{i}}\varphi(x)| = o(P(x)), |x| \to \infty,$$
for all $\mathbf{i} = (i_1, i_2, \cdots, i_d), 0 \le |\mathbf{i}| \le 6$.

Core Lemmas

1. **Regularity of the PDE solution:** Backward Kolmogorov Equation:
$$\begin{cases} \dfrac{\partial u}{\partial t} + \mathcal{L}u = 0, \\ u(x,\tau) = \varphi(x) \end{cases}$$
where
$$\mathcal{L} := a(x)\cdot\nabla + \frac{\sigma^2}{2}\Delta.$$
the Kolmogorov solution u is smooth and its spatial derivatives decay exponentially as $t \to \tau$, which controls higher-order terms.
2. **Talay–Tubaro local expansion:** expanding one-step increments identifies the local weak error coefficient B_0, yielding
$$R = C_0(\tau,x)h + O(h^2).$$
3. **Moment bounds for the numerical scheme:**
$$\mathbf{E}|X_k|^{2p} \le C\left(1 + |x|^{2p}e^{-\gamma t_k}\right),$$
4. **Vanishing invariant average of B_0:** under the Gibbs invariant density ρ,
$$\int B_0(t,y)\rho(y)\,dy = 0,$$
so the first-order error cancels at equilibrium.

Summary and Discussion

Many SDE cases offer opportunities to develop and deploy specialised methods suited to the specific problem. In this paper, authors have furthered our understanding of one such, non-Markovian method, that offers promising numerical efficiently improvements for certain cases of stochastic gradient systems. We hope further work can test the scheme in more cases and prove more bounds on the constant in the scheme's $O(h^2)$ convergence.

References

1. Leimkuhler, B., Matthews, C., Tretyakov, MV. 2014, On the long-time integration of stochastic gradient systems. Proc.R.Soc.A470: 20140120. http://dx.doi.org/10.1098/rspa.2014.0120.

Main Theorem

Theorem

Assume that all assumptions hold. For the non-Markovian scheme

$$X_{k+1} = X_k + h\,a(X_k) + \sigma\frac{\sqrt{h}}{2}(\xi_k + \xi_{k+1}),$$

the weak error admits the expansion

$$R = \mathbf{E}\varphi(X^x(\tau)) - \mathbf{E}\varphi(X_N) = C_0(\tau,x)h + C(\tau,x)h^2,$$

where

$$C_0(\tau,x) = \mathbf{E}\int_0^\tau B_0(t, X^x(t))dt$$

with the leading coefficient B_0 in local weak error expansion. Moreover, the expectation of leading coefficient satisfies following relation with constants $\ell, \lambda > 0$

$$|C_0(\tau,x)| \le K(1+|x|^\ell)e^{-\lambda\tau}.$$

Hence

$$C_0(\tau,x) \to 0, \tau \to \infty.$$

Therefore, although the scheme is only first-order weakly accurate on finite time intervals,

$$\mathbf{E}\varphi(X^x(\tau)) - \mathbf{E}\varphi(X_N) = O(h),$$

it becomes second-order accurate in the long-time ergodic regime:

$$\lim_{\tau\to\infty}|\mathbf{E}\varphi(X^x(\tau)) - \mathbf{E}\varphi(X_N)| = O(h^2).$$

Proof Strategy

The proof consists of two connected parts: a finite-time weak error expansion and a long-time cancellation argument.

1. **Rewrite the global weak error as a telescoping sum.** Using the backward Kolmogorov solution $u(t,x)$, write
$$R = \sum_{k=0}^{N-1}\mathbf{E}\left[u(t_{k+1}, X^{t_k,X_k}(t_{k+1})) - u(t_{k+1}, X_{k+1})\right].$$
2. **Integral representation:** Define
$$C_0(\tau,x) = \mathbf{E}\int_0^\tau B_0(t, X^x(t))\,dt,$$
giving
$$R = C_0(\tau,x)h + O(h^2).$$
with remaining leading contribution
$$B_0(t,x) = \frac{1}{2}\left[\sum_{i,j} a_j\frac{\partial a_i}{\partial x_j}\frac{\partial u}{\partial x_i} + \frac{\sigma^2}{2}\sum_{i,j}\frac{\partial a_i}{\partial x_j}\frac{\partial^2 u}{\partial x_i \partial x_j} + \frac{\sigma^2}{2}\sum_{i,j}\frac{\partial^2 a_i}{(\partial x_j)^2}\frac{\partial u}{\partial x_i}\right].$$
and by **Lemma 1,2,3**, induce
$$|C_0(\tau,x)| \le K(1+|x|^\ell)e^{-\lambda\tau}.$$
3. **Prove that $C_0(\tau,x)$ decays to zero.** Write
$$C_0(\tau,x) = \int_0^\tau\int_{\mathbb{R}^d} B_0(t,y)p(t,x,y)\,dy\,dt.$$
Insert and subtract the invariant density ρ:
$$p(t,x,y) = \rho(y) + \left[p(t,x,y) - \rho(y)\right].$$
The invariant part vanishes by **Lemma 4** while the remaining part is bounded by exponential ergodicity. Therefore
$$|C_0(\tau,x)| \le K(1+|x|^\ell)e^{-\lambda\tau}.$$

Numerical Experiments

After proving the main theorems, there are two major practical concerns for the authors try to test with numerical experiments.

- Can we observe the theoretical exponential convergence to $O(h^2)$ error of distribution within a reasonable time frame?
- Does the method provide practical efficiency improvements for the relevant type of problem?

To test convergence, authors devise a toy problem with $V(x) = cos(x)$, resulting in SDE: $dX = \sin(X)dt + \sigma dW$.
Three methods are compared on the problem,

- A simple $O(h)$ Euler–Maruyama scheme.
- A Heun's method scheme, known to be known $O(h^2)$ for this problem, requiring two computationally expensive evaluations of ∇V per time step.
- The studied new candidate scheme.

For a sufficiently small time step, L_2 errors of the distribution produced by the toy problem SDE behave over time as in figure 1, with exponential decay of one term in the candidate scheme error as $t \to \infty$. We can clearly see the decay of the C_0, the $O(h)$ term, with the candidate scheme even outperforming the Heun's method for a given time step once sufficient time has passed.

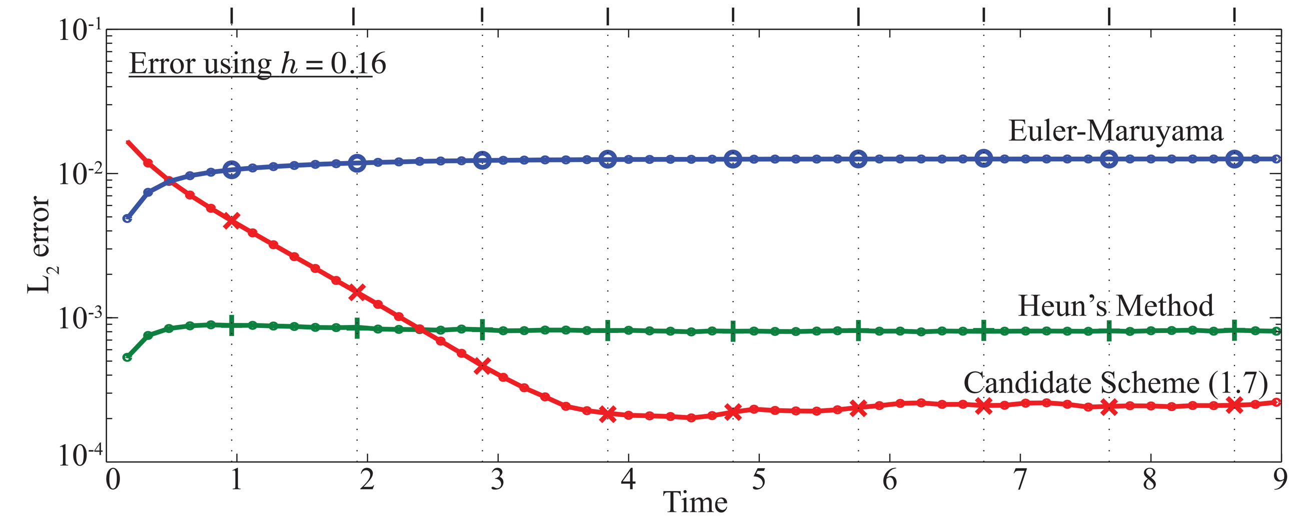

Figure: L_2 error of the distribution of the toy problem over time for the three methods using small time step $h = 0.16$

Next, because the candidate scheme discussed by this paper offers potential efficiency improvements especially for high dimensional potentials, where determining $-\nabla V$ dominates the computational load in each time step. Authors test such a case with a 192-dimensional Lennard-Jones box,

$$V(q) = \sum_{i=1}^{64}\sum_{j=i+1}^{64}\left(r_{ij}^{-12} - r_{ij}^{-6}\right), \qquad r_{ij} = \|q_i - q_j\|.$$

Using the same three schemes, authors track error convergence as $h \to 0$ for a large enough t. Results, shown in figure 2, show both good agreement with the theoretical results, and the candidate scheme having a small observed error constant for its Ch^2 convergence, when compared to the Heun's method. With the studied scheme also being nearly twice as fast as the Heun's method for this problem, this example shows a promising use case for the candidate scheme.

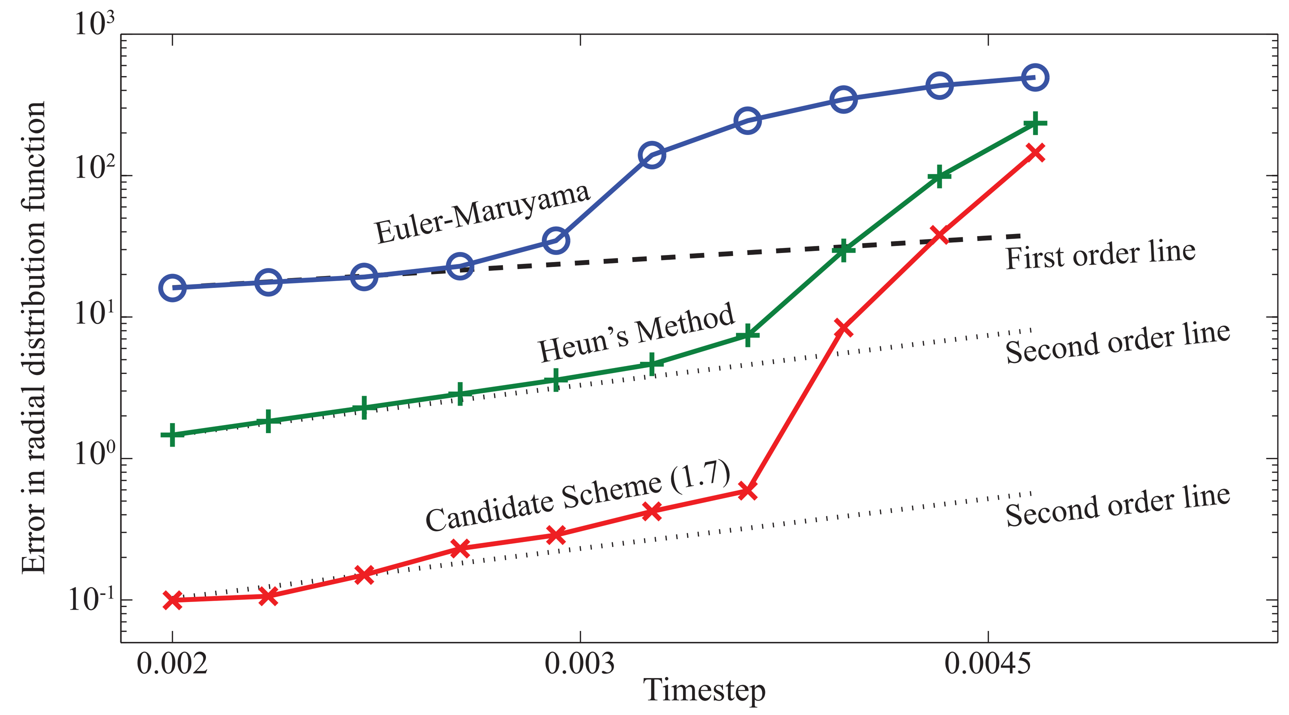

Figure: Convergence in distribution error for a high dimensional Lennard-Jones box using the three methods.