Adaptive stepsize algorithms for Langevin dynamics

Alix Leroy[1,*]

Joint work with supervisors: Ben Leimkuhler [1], Jonas Latz[2], Des Higham [1]

[1]School of Mathematics, University of Edinburgh [2] Department of Mathematics, University of Manchester

*s2133976@ed.ac.uk.

[a]Supported by the Luxembourg National Research Fund (FNR) (16978012)

1. Background

Goal: Sampling via the invariant distribution $\rho_\infty(\mathbf{x}) \propto \exp(-\beta \mathbf{V}(\mathbf{x}))$ of stochastic differential equation.

Using Langevin dynamics in the large friction limit, we get the overdamped stochastic differential equation:

$$\mathrm{d}X = -\nabla V(X)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}W(t)$$

where $W(t)$ is the Brownian motion and $V(x)$ is a potential. With a numerical scheme, samples from the invariant distribution $\rho_\infty(x) = C\exp(-\beta V(x))$ can be drawn integrating one path for a long time. Euler-Maruyama applied to the overdamped SDE yields:

$$X_{i+1} = X_i - \nabla V(X_i)h + \sqrt{2\beta^{-1}h}\Delta W_i,$$

the time is discretised with $h = 1/n_s$, n_s is the number of samples, $X_i \approx X(t = ih)$ and $\Delta W_i \sim \mathcal{N}(0,1)$.

2. Motivating example

Sampling from a modified harmonic potential with a sudden change of frequency:

$$V'(x) = (\omega(x)^2 + c)x$$
$$\omega(x) = \frac{b}{\frac{b}{a} + (x - x_0)^2}$$

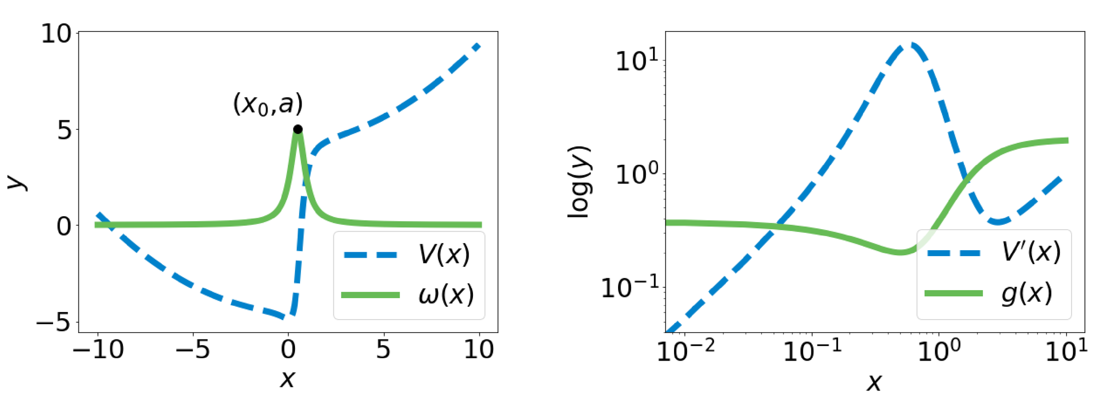

Figure 1: On the left, the modified harmonic potential $V(x)$ and the function $\omega(x)$ with $a = 5$, $x_0 = 0.5$ and $b = 1$. A smaller b implies a steeper $\omega(x)$ around x_0. On the right, the gradient $V'(x)$ and an example of adaptive function $g(x)$.

To reduce the computational cost, the size of the steps could be smaller around singularities and bigger elsewhere driven by a function $g(x)$, for instance:

$$g(x) \propto \frac{1}{\|V'(x)\|},$$

with $g(x) \neq 0$ and $g(x) \in [\frac{\Delta t_{\min}}{\Delta t}, \frac{\Delta t_{\max}}{\Delta t}]$.

Naive time re-scaling in the continuous dynamics:

$$dt \longrightarrow g(x)\mathrm{d}\tau$$
$$dW(t) \longrightarrow \sqrt{g(x)}\mathrm{d}W(\tau)$$

gives the naive time rescaled SDE:

$$\mathrm{d}\tilde{X} = -\nabla V(\tilde{X})g(\tilde{X})\mathrm{d}\tau + \sqrt{2\beta^{-1}g(\tilde{X})}\mathrm{d}W(\tau).$$

Euler-Maruyama scheme applied to this SDE:

$$\tilde{X}_{i+1} = \tilde{X}_i - \nabla V(\tilde{X}_i)g(\tilde{X}_i)h + \sqrt{2\beta^{-1}g(\tilde{X}_i)h}\Delta W_i.$$

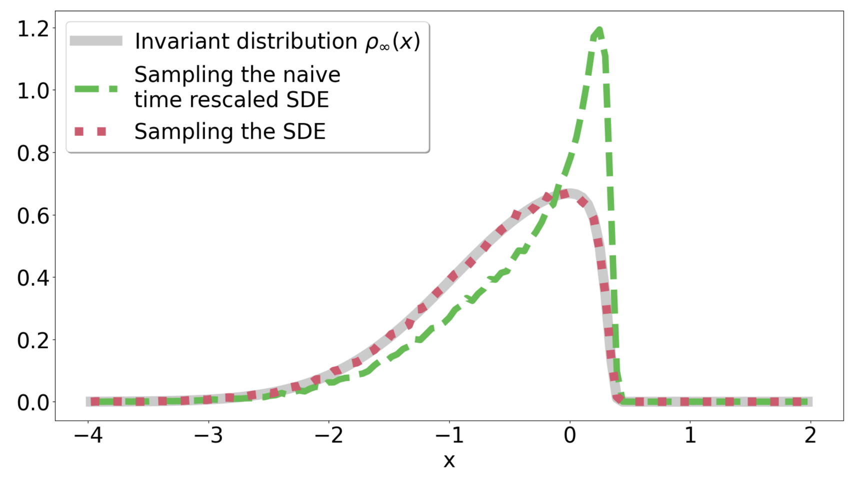

Figure 2: Histograms of the invariant distribution and samples from Euler-Maruyama scheme applied to the original SDE and to the naive time rescaled SDE after $50\,000$ steps, $h = 0.001$, $n_s = 10^5$, $\frac{\Delta t_{\max}}{\Delta t} = 2$, $\frac{\Delta t_{\min}}{\Delta t} = 0.001$, $\tau = 0.1$ ($c = 0.1, b = 0.1, a = 10, x_0 = 0.5$).

In fact, using the backward Kolmogorov equation, one can show that $\rho_\infty(x)$ is not the invariant distribution for the naive time rescaled SDE.

3. An IP-transformed SDE

The bias in the invariant distribution arises from the factor $\sqrt{g(x)}$ in the diffusion term which gives rise to an increasing variance of displacement. The invariant preserving (IP) -transformed SDE leading to the appropriate distribution is:

$$\mathrm{d}Y = -\nabla V(Y)g(Y)\mathrm{d}s + \beta^{-1}\nabla g(Y)\mathrm{d}s + \sqrt{2\beta^{-1}g(Y)}\mathrm{d}W_s.$$

We can verify using the backward Kolmogorov equation that $\rho_\infty(x)$ is the invariant distribution of this direct time-rescaled transformed SDE. Euler-Maruyama on this SDE:

$$Y_{i+1} = Y_i - \nabla V(Y_i)g(Y_i)h + \beta^{-1}\nabla g(Y_i)h + \sqrt{2\beta^{-1}g(Y_i)h}\Delta W_i.$$

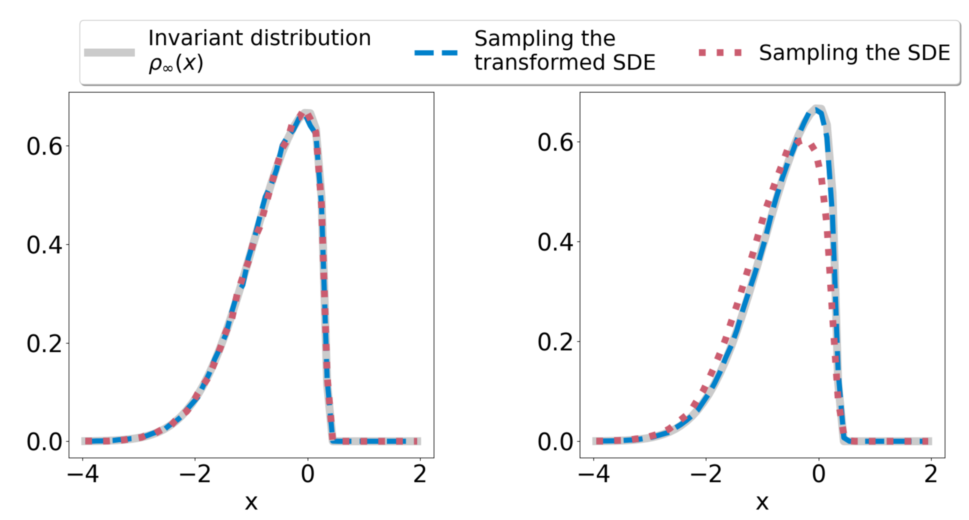

Figure 3: On the left, histograms with a small steps $h = 0.001$, both SDEs converge. On the right histograms with a larger stepsize $h = 0.05$, only the transformed SDE converges. Simulations use similar parameters to figure 2.

For a large stepsize, the IP-transformed SDE converges toward the invariant distribution while the overdamped SDE fails to converge. In Figure 5, the rescaling factor is **1.18** but the error is significantly smaller for the IP-transformed SDE.

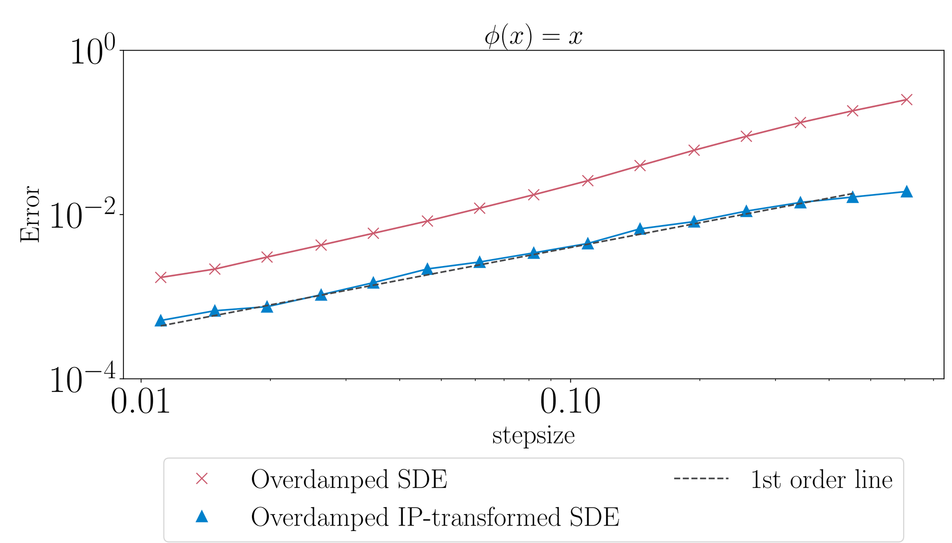

Figure 4: The order of convergence of the Euler-Maruyama method is recovered for larger stepsize for both dynamics, but the error is significantly smaller with the IP-transformed SDE.

4. IP-transformed underdamped

A separable Hamiltonian dynamics, $H(q,p) = E_{\text{kin}}(p) + V(q)$, with position $q \in \mathbb{R}^d$ and momentum $p \in \mathbb{R}^d$, with the friction γ has the corresponding IP-transformed dynamic:

$$\begin{cases} dq &= pg(q)ds, \\ \mathrm{d}p &= -\nabla V(q)g(q)ds + \beta^{-1}\nabla g(q)ds - \gamma p g(q)ds + \\ & \sqrt{2\gamma\beta^{-1}g(q)}dW(s). \end{cases}$$

with invariant distribution is $\rho(q,p) \propto \exp(-\beta H(q,p))$.

We use numerical timestepping integration methods based on splitting schemes, which break the equations into separate parts to be solved independently

$$\begin{bmatrix} dq \\ dp \end{bmatrix} = \underbrace{\begin{bmatrix} g(q)p \\ 0 \end{bmatrix}}_{A} dt + \underbrace{\begin{bmatrix} 0 \\ -\nabla V(q)g(q) \end{bmatrix}}_{B} dt + \underbrace{\begin{bmatrix} 0 \\ \beta^{-1}\nabla g(q)dt - \gamma p g(q)dt + \sqrt{2\beta^{-1}g(q)}dW(t) \end{bmatrix}}_{O}.$$

The IP-transformed system of SDEs requires the computation of extra terms such as $\Delta g(x)\beta^{-1}$ and renders the step A implicit. We design novel schemes based on modified building blocks such as $\hat{B}\hat{A}\hat{O}\hat{A}\hat{B}$ and $\tilde{B}\tilde{A}\tilde{O}\tilde{A}\tilde{B}$.

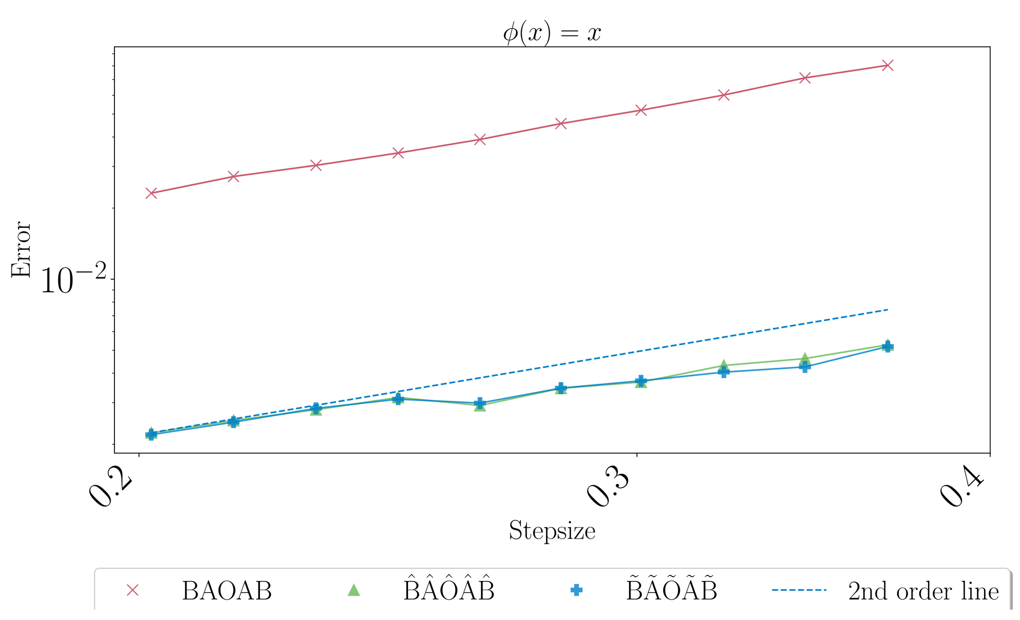

Figure 5: The schemes $\tilde{B}\tilde{A}\tilde{O}\tilde{A}\tilde{B}$ and $\hat{B}\hat{A}\hat{O}\hat{A}\hat{B}$ have much higher efficiency and display a quadratic decay in the error, similarly to the original schemes BAOAB.

5. Conclusions

- Invariant-preserving transformed dynamics are based on an embedded adaptive time stepping based on a monitor function.
- We provide several novel splitting schemes to integrate the invariant-preserving transformed underdamped system.
- Higher efficiency is shown for the numerical integration of the IP-transformed dynamics.

- **Preprint with detailed analytical considerations**