Introduction to Ergodicity and Numerical Analysis for Langevin Dynamics

Aaron Kirtland

2021/5/3

Langevin Dynamics Introduction

Hamiltonian dynamics gives

$$\dot{q} = \frac{\partial H}{\partial p} \qquad \dot{p} = -\frac{\partial H}{\partial q}$$

In molecular dynamics, usually $H = p^T M^{-1} p/2 + U(q)$ for M a diagonal mass matrix. So the dynamics read

$$\dot{q} = M^{-1} p \qquad \dot{p} = -\nabla U(q)$$

However, systems we usually want to consider don't exist in a vacuum; they experience frictional forces and their statistical behavior changes with temperature.

Langevin Dynamics Equation

$$
\begin{aligned}
\mathrm{d}q &= M^{-1} p \,\mathrm{d}t \\
\mathrm{d}p &= -\nabla U(q)\,\mathrm{d}t - \gamma p\,\mathrm{d}t + \sqrt{2\gamma k_B T} M^{1/2}\,\mathrm{d}W
\end{aligned}
$$

where

- $q, p \in \mathbb{R}^d$, W is d-dimensional Brownian motion, $U : \mathbb{R}^d \to \mathbb{R}$
- M represents the masses of a system of particles
- U is the potential energy
- k_B is Boltzmann's constant
- T is temperature
- γ is the friction coefficient or collision rate, with units $1/$time

Relationships with Other Common Dynamics Equations

- We can derive Brownian dynamics as the overdamped limit of Langevin dynamics where we let $v = M^{-1}p$, assume $\mathrm{d}p / \mathrm{d}t = 0$, and solve for q:

$$\mathrm{d}q = -\gamma^{-1} M^{-1} \nabla U(q)\,\mathrm{d}t + \sqrt{2k_B\gamma^{-1}T} M^{-1/2}\,\mathrm{d}W$$

- We can decompose Langevin dynamics as

$$\mathrm{d}\begin{pmatrix} q \\ p \end{pmatrix} = \underbrace{\begin{pmatrix} M^{-1}p \\ 0 \end{pmatrix}\mathrm{d}t}_{\text{A}} + \underbrace{\begin{pmatrix} 0 \\ -\nabla U(q) \end{pmatrix}\mathrm{d}t}_{\text{B}} + \underbrace{\begin{pmatrix} 0 \\ -\gamma p\,\mathrm{d}t + \sigma M^{1/2}\,\mathrm{d}\omega \end{pmatrix}}_{\text{O}}$$

O gives the Ornstein-Uhlenbeck equation.

The Generator

For $\phi \in C^2$ a suitable function that decays to 0 exponentially with x, Itô's rule states

$$\mathrm{d}\phi(x) = \phi'(x)(f(X)\,\mathrm{d}t + g(x)\,\mathrm{d}W) + \frac{1}{2}\phi''(x)g(x)^2\,\mathrm{d}t$$

Then the expectation satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(\mathrm{d}\phi(X)) = \mathbb{E}((\mathcal{L}\phi)(X)) = \int_{-\infty}^{\infty}(\mathcal{L}\phi)(x)\rho(x,t)\,\mathrm{d}x$$

where $\mathcal{L} = f(x)\frac{\partial}{\partial x} + \frac{1}{2}g(x)^2\frac{\partial^2}{\partial x^2}$ is a linear operator, the generator of the stochastic process.

The Kolmogorov Operator

From the proof of the Fokker-Planck equation, we have

$$\langle \mathcal{L}\phi, \rho \rangle = \langle \phi, \mathcal{L}^\dagger \rho \rangle = \int_{-\infty}^{\infty} (-\frac{\partial}{\partial x}(f(x)\rho(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(g(x)^2\rho\right))\varphi \, \mathrm{d}x$$

where $\mathcal{L}^\dagger$, the Kolmogorov operator, is the adjoint of the generator operator and when acting on ρ gives the right hand side of the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^\dagger \rho$$

If $g(x) = 0$, then $\mathcal{L}$ and $\mathcal{L}^\dagger$ coincide with the Lie derivative and the Liouvillian, respectively.

Drift and Minorization Method

Let $\mathrm{d}X = Y\mathrm{d}t + \sum_i X_i\,\mathrm{d}\omega_i$. Let $P_t(x, A) = \mathbb{P}(x(t) \in A | x(0) = x)$.

1. (Drift) There exists a radially unbounded Lyapunov function satisfying for some constants $\alpha, \delta > 0$ and all $x \in D$

$$\mathcal{L}\varphi \leq -\alpha\varphi + \delta$$

   Then let $C = \{x \in D : \varphi(x) \leq 2\beta/(\gamma - \alpha)\}$.

2. (Minorization) There exists a sampling rate T, $\eta > 0$, and a probability measure ν on C such that $P_T(x, A) \geq \eta\nu(A)$.

Then there exists a unique invariant measure p_* with $\mathcal{L}^\dagger p_* = 0$ and there exist k, λ such that for all suitable f with $|f| \leq \varphi$,

$$|(\exp(t\mathcal{L})f)X_0 - \int_D f(z)p_*(z)\,\mathrm{d}z| \leq k\exp(-\lambda t)\varphi(x_0)$$

Proving Minorization

Equivalent to

1. There exists a $y \in \operatorname{int} C$ such that for all $\delta > 0$, there exists a t with $P_t(x, B_\delta(y)) > 0$.
2. For all t, there exists a C^0 density $p_t(x, y)$, i.e. for all $x \in C$, $\mathcal{B}(\mathbb{R}^n) \cap \mathcal{B}(C)$,

$$P_t(x, A) = \int_A p_t(x, y)\,\mathrm{d}y$$

By Hörmander's Theorem, the second assumption is equivalent to Hörmander's condition, which we used previously in the context of Chow's theorem for local controllability.
Hörmander's Condition: the ideal generated by $\{X_1, \ldots, X_m\}$ in $\{Y, X_1, \ldots, X_m\}_{LA}$ spans $\mathbb{R}^n$ for all x

Application to Langevin Dynamics

- For Langevin dynamics, if $M = I$, we can use the Lyapunov function(s) $H^l(q,p) = \left(\frac{1}{2}||p||^2 + U(q)\right)^l$ for l a positive integer.
- We want l to vary so that the condition $|f| \leq H^l$ can hold for many observables f.

Let $Y = \begin{pmatrix} p \\ -\gamma p - \nabla U(q) \end{pmatrix}$ and $\sqrt{2\gamma k_B T} M^{1/2}\,\mathrm{d}W = \sum_{i=1}^{d} X_i\,\mathrm{d}W_i$, so $X_i = (0, \rho_i)^T$ for $\rho_i \in \mathbb{R}^d$. Then for Hörmander's Condition, we see

$$[X_i, Y] = (DY, X_i) = \begin{pmatrix} 0 & I \\ -d^2U(q) & -\gamma I \end{pmatrix} \begin{pmatrix} 0 \\ \rho_i \end{pmatrix} = \begin{pmatrix} \rho_i \\ -\gamma \rho_i \end{pmatrix}$$

so $\{X_1, \ldots, X_d, [X_1, Y], \ldots, [X_d, Y]\}$ spans $\mathbb{R}^{2d}$.

Canonical Distribution

- The canonical NVT ensemble is a natural system to study for molecular dynamics
- Let $\rho_\beta(q,p) = Z^{-1}\exp\big(-\beta H^1(q,p)\big)$, the Boltzmann/canonical distribution, where Z is the canonical partition function $Z = \int_D \exp(-\beta H(q,p))\,\mathrm{d}x$.
- We can verify $\mathcal{L}_{LD}^\dagger \rho_\beta = 0$, so this is the measure with respect to which Langevin dynamics are ergodic.
- Therefore, we can study the canonical distribution with Langevin dynamics using different observables f

Splitting Methods

Recall that we can decompose Langevin dynamics as

$$\mathrm{d}\begin{pmatrix} q \\ p \end{pmatrix} = \underbrace{\begin{pmatrix} M^{-1}p \\ 0 \end{pmatrix}\mathrm{d}t}_{\text{A}} + \underbrace{\begin{pmatrix} 0 \\ -\nabla U(q) \end{pmatrix}\mathrm{d}t}_{\text{B}} + \underbrace{\begin{pmatrix} 0 \\ -\gamma p\,\mathrm{d}t + \sigma M^{1/2}\,\mathrm{d}\omega \end{pmatrix}}_{\text{O}}$$

This yields updates individually as

$$\begin{aligned} \mathcal{U}_h^A(q,p) &= (q + hM^{-1}p, p) \\ \mathcal{U}_h^B(q,p) &= (q, p - h\nabla U(q)) \\ \mathcal{U}_h^O(q,p) &= (q, e^{-\gamma h}p + \sqrt{k_B T(1 - e^{-2\gamma h})}M^{1/2}R) \end{aligned}$$

where R is a vector of d i.i.d. normal random numbers.

Splitting Methods Continued

We define splitting methods in the following fashion:

$$\mathcal{U}_h^{[\![\mathrm{BABO}]\!]} = \mathcal{U}_h^{\mathrm{O}} \mathcal{U}_{h/2}^{\mathrm{B}} \mathcal{U}_h^{\mathrm{A}} \mathcal{U}_{h/2}^{\mathrm{B}}$$

Different splitting schemes produce different results, for example consider the 1D harmonic oscillator with spring constant Ω^2, so $U(q) = \Omega^2 q^2/2$, $\gamma = 0$. We find the long term averages satisfy

Scheme	$\langle q^2 \rangle_h / \langle q^2 \rangle$	$\langle p^2 \rangle_h / \langle p^2 \rangle$	$\langle qp \rangle_h$
$[\![\mathrm{ABOBA}]\!]$	1	$(1 - h^2\Omega^2/4m)^{-1}$	0
$[\![\mathrm{OABAO}]\!]$	$1 - h^2\Omega^2/4m$	1	0
$[\![\mathrm{BAOA}]\!]$	1	1	$O(h)$

Correction Functions

The averages we compute $\langle \varphi \rangle_h$ of observables are only approximations of $\langle \varphi \rangle$. How can we analyze the accuracy of these values?

Suppose $\hat{\rho} = \rho_\beta(1 + hf_1 + h^2 f_2 + h^3 f_3 + O(h^4))$ where $f_k(q,p)$ are correction functions satisfying $\int_D f_k \rho_\beta \,\mathrm{d}x = 0$ so that the partition function is preserved, $\int_D \hat{\rho}\,\mathrm{d}x = \int_D \rho_\beta \,\mathrm{d}x$.

$$
\begin{aligned}
\langle \varphi \rangle_h &= \int_D \varphi \hat{\rho}\,\mathrm{d}x \\
&= \int_D \varphi \rho_\beta \,\mathrm{d}x + h \int_D \varphi f \rho_\beta + O(h^2) \\
&= \langle \varphi \rangle + h \langle \varphi f_1 \rangle + h^2 \langle \varphi f_2 \rangle + O(h^3)
\end{aligned}
$$

Therefore, we would like to compute the correction functions f_i.

Computing f_i: First Try

Now let's assume that $\hat{\mathcal{L}}^\dagger = \mathcal{L}^\dagger_{\mathrm{LD}} + h\mathcal{L}^\dagger_1 + h^2\mathcal{L}^\dagger_2 + O(h^3)$ for some perturbation operators $\mathcal{L}^\dagger_i$. Then in solving for a stationary solution to the Fokker-Planck equation, we have $\hat{\mathcal{L}}^\dagger\hat{\rho} = 0$, or

$$\left(\mathcal{L}^\dagger_{\mathrm{LD}} + h\mathcal{L}^\dagger_1 + h^2\mathcal{L}^\dagger_2 + O(h^3)\right)\left(\rho_\beta(1 + hf_1 + h^2 f_2 + h^3 f_3 + O(h^4))\right) = 0$$

from which we obtain the first-order approximation $\mathcal{L}^\dagger_{\mathrm{LD}}(\rho_\beta f_1) + \mathcal{L}^\dagger_1\rho_\beta = 0$ as $\mathcal{L}^\dagger_{\mathrm{LD}}\rho_\beta = 0$. If we know $\mathcal{L}^\dagger_1$, then we can solve this at least numerically for f_1.

Application to $[\![OBA]\!]$

$$\exp\left(h\hat{\mathcal{L}}^\dagger_{[\![OBA]\!]}\right) = \exp\left(h\mathcal{L}^\dagger_A\right)\exp\left(h\mathcal{L}^\dagger_B\right)\exp\left(h\mathcal{L}^\dagger_O\right)$$

$$\mathcal{L}^\dagger_A = -p\frac{\partial}{\partial q} \quad \mathcal{L}^\dagger_B = \nabla U(q)\frac{\partial}{\partial p} \quad \mathcal{L}^\dagger_O = \gamma(\mathrm{Id} + p\frac{\partial}{\partial p}) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial p^2}$$

The BCH formula gives

$$\hat{\mathcal{L}}^\dagger_{[\![OBA]\!]} = \mathcal{L}^\dagger_A + \mathcal{L}^\dagger_B + \mathcal{L}^\dagger_O + \frac{h}{2}\left([\mathcal{L}^\dagger_A, \mathcal{L}^\dagger_B] + [\mathcal{L}^\dagger_A, \mathcal{L}^\dagger_O] + [\mathcal{L}^\dagger_B, \mathcal{L}^\dagger_O]\right) + O(h^2)$$

where $[\cdot,\cdot]$ is the commutator bracket.

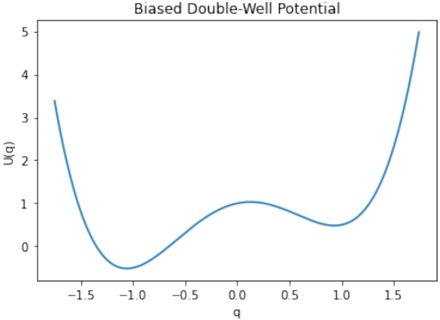
$[\![\text{OBA}]\!]$ Continued

Then letting the first order approximation of this be $\hat{\mathcal{L}}^\dagger$, we find that $[\![\text{OBA}]\!]$ behaves like it's solving the PDE $\rho_t = \hat{\mathcal{L}}^\dagger \rho$ where

$$\hat{\mathcal{L}}^\dagger \rho = \mathcal{L}^\dagger_{\text{LD}} \rho + \frac{h}{2}\left(-pU''(q)\rho_q + U'(q)\rho_q + \gamma p \rho_q + \sigma^2 \rho_{qp} + \gamma U'(q)\rho_p\right) + O(h^2)$$

Therefore, this scheme is amenable to the correction function method!

Numerical Simulation



- Define the biased double-well potential $U(q) = (q^2 - 1)^2 + q/2$.
- Let $M = \beta = 1$.
- Let $v(q, p) = p^2 - qU'(q) + 2qp$.
- Then $\langle v \rangle = 0$.

Results

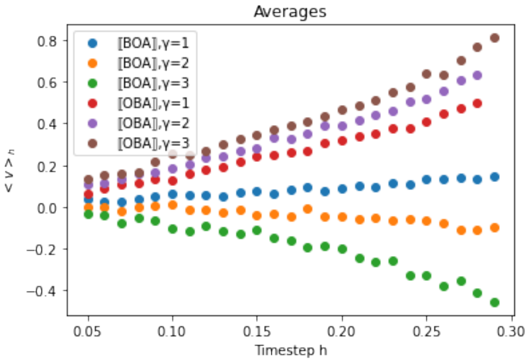


Figure: Averages over 3 Trials, each over a total time of 10000 units. Observe that for $\gamma = 2$, the $[\![$BOA$]\!]$ scheme has very low error.

$[\![BOA]\!]$ Scheme Result Explanation

We can explain this rigorously using the correction function method because

$$\langle vf_{1,[\![BOA]\!]}\rangle = \frac{\beta}{2}\int_{\mathbb{R}^2} v(\gamma U(q) + pU'(q) - c)\rho_\beta(q,p)\,\mathrm{d}x = \frac{2-\gamma}{2\beta}$$

where $c = \gamma\langle U(q)\rangle$.

So the first order correction is 0 for $\gamma = 2$, hence we are observing accurate results up to second order! Therefore, in practice the choice of method can depend on the value of γ.

Conclusion

- Langevin dynamics arises naturally from physical considerations.
- It is geometrically ergodic.
- We can find various splitting methods for its SDE.
- These splitting methods can be analyzed with correction functions.
- The proofs of these facts use many of the same tools we studied this semester.

Questions

Thanks for listening! Any questions?

Sources:

- Leimkuhler, B., & Matthews, C. (2016). *Molecular Dynamics*. Springer International PU.
- Mattingly, J. C., Stuart, A. M., & Higham, D. J. (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic processes and their applications, 101(2)*, 185-232.