ESE 559 Final Report: Introduction to Ergodicity and Numerical Analysis for Langevin Dynamics

Aaron Kirtland

2021/5/3

1 Abstract

Langevin dynamics are commonly used in molecular dynamics to represent a collection of particles in a medium, and therefore efficient numerical methods for computing them are of the utmost importance. In this work, we introduce Langevin dynamics and relevant concepts in stochastic differential equations (SDEs) and chemistry, provide a basic theorem for geometric ergodicity of SDEs and apply it to Langevin dynamics, and discuss and analyze a family of splitting methods for numerically solving the Langevin equations. We supplement this discussion with a numerical experiment of simulating a biased double well system with two different splitting schemes. The results we present in this paper were strongly influenced by their presentation in [1], though we supplement the results in this report with results from revelant papers like [2].

2 Background

2.1 Langevin Dynamics

To introduce Langevin dynamics, we begin with recalling Hamiltonian dynamics. For $p, d \in \mathbb{R}^d$, these state that

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

where H is the Hamiltonian of the system of interest, q is the vector of positions, and p is the vector of momentum. In molecular dynamics, usually $H = p^T M^{-1} p/2 + U(q)$ for M a diagonal matrix containing the

masses of each particle of interest. In this case, the dynamics read

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

However, while Hamiltonian dynamics assumes that the system is closed and that the particles described by the system are the only particles contained in the system, it is often impossible to describe system of interest in this manner, partially because of the sheer number of particles in systems of interest. Equivalently, the systems we usually want to consider don't exist in a vacuum; the experience other forces like friction, and their statistical behavior appears to change randomly in a way related to temperature. This leads us to consider Langevin dynamics, where we add frictional and white noise terms to Hamiltonian dynamics. In particular, they state

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

where W is d-dimensional Brownian motion, $U : \mathbb{R}^d \to \mathbb{R}$ is the potential energy, k_B is Boltzmann's constant, T is the temperature, and γ is the friction coefficient or collision rate with units 1/time.

From Langevin dynamics, we can derive several other systems of physical significance. If we take the overdamped limit of Langevin dynamics, meaning we set $v = M^{-1}p$, assume $\mathrm{d}p \,/\, \mathrm{d}t = 0$, and solve for q, we derive Brownian dynamics to be

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

We also notice that we can split Langevin dynamics as

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

where if we consider only the A and B parts we obtain Hamiltonian dynamics, and if we consider only the O part, we obtain the Ornstein-Uhlenbeck equation.

2.2 SDEs

Let $\mathrm{d}x = f(x,t)\,\mathrm{d}t + g(x,t)\,\mathrm{d}W$. Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be continuously twice differentiable. Then Itô's rule states

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

Now let's assume in particular that $f, g, \varphi : \mathbb{R}^n \to \mathbb{R}^n$. Define C_P^∞ as the set of functions $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ such that $\varphi(x) = O(|x|^m)$ for some positive integer m as $|x| \to \infty$. Also, let's assume that the density ρ of x is C^∞ for all t and that it decays exponentially to 0 as $|x| \to \infty$ and also that $f, g, \varphi \in C_P^\infty$. Then the Fokker-Planck equation for the system, and by Itô's rule the expectation of $\varphi(x)$ 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. Additionally, from the proof of the Fokker-Planck equation, we have

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

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

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

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

2.3 The Canonical Distribution

Chemists and physicists try to consider the probability of observing particles defined by a state vector (q,p) in regions of the state space, and they also try to answer how observables, functions of the state space (q,p), behave in terms of averages over regions of the state space. For this reason, it is natural to define a measure on the state space D, and statistical mechanics gives us a natural measure, the canonical measure, to define in the case where the number of particles N, the volume V, and the temperature T is fixed. This applies to many systems of interest in molecular dynamics, and this framework is called the canonical ensemble.

We define $\rho_\beta(q,p) = Z^{-1}\exp\big(-\beta H^1(q,p)\big)$ to be this canonical or Boltzmann distribution, where Z

is the partition function $Z = \int_D \exp(-\beta H(q,p))\,\mathrm{d}x$ that gives a normalization constant, and $\beta = 1/k_b T$. Langevin dynamics is defined in such a way such that it has a unique stationary distribution, and we can verify $\mathcal{L}_{\mathrm{LD}}^\dagger \rho_\beta = 0$, so we can use Langevin dynamics to study the canonical ensemble.

3 Ergodicity

One property that we would like to have for practical simulation purposes is that the long-time average of an observable over trajectories beginning from most initial conditions converges to the average provided by the canonical distribution. This property holds in a certain precise way for Langevin dynamics through geometric ergodicity.

A general way of proving geometric ergodicity for a variety of SDEs is the drift-and-minorization condition popularized through works like ([3],[4]). Here we will state one version which is influenced by both [1] and [2].

Let $D \subset \mathbb{R}^n$ be open. Define the $\mathrm{d}x = Y(x)\,\mathrm{d}t + \sum_i X_i(x)\,\mathrm{d}W_i$ on D. Let $P_t(x, A) = \mathbb{P}(x(t) \in A | x(0) = x)$ where $A \in \mathcal{B}(D)$ the set of Borel sets on D. Suppose that the following conditions hold:

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 we define a compact set $C = \{x \in D : \varphi(x) \leq K_{\alpha,\delta}\}$ for K a constant chosen that depends on α and δ that allows the theorem to hold.

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 the system is geometrically ergodic, meaning there exists a unique invariant measure p_* with $\mathcal{L}^\dagger p_* = 0$ and there exist $k, \lambda > 0$ such that for all suitable f with $|f| \leq \varphi$,

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

In words, this states that the distance between f evaluated along the trajectory and the average of f around the region under the invariant measure is bounded by a decreasing exponential function of time, which coincides with our previously stated concept of ergodicity.

The minorization condition is equivalent to the following two conditions:

1. There exists a $y \in \operatorname{int} C$ such that for all $\delta > 0$, there exists a t with $P_t(x, B_\delta(y)) > 0$, or in other words for any neighborhood of y, there exists some time t where the system trajectory beginning at x hits the neighborhood with positive probability.

2. For all t, there exists a continuous density $p_t(x, y)$, i.e. for all $x \in C$, $\mathcal{B}(D) \cap \mathcal{B}(C)$,

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

By Hörmander's Theorem, the second assumption is equivalent to Hörmander's condition, which in this case states that the ideal generated by $\{X_1, \ldots, X_m\}$ in $\{Y, X_1, \ldots, X_m\}_{LA}$ spans $\mathbb{R}^n$ for all $x \in D$. We note that Hörmander's condition can also be used to prove local controllability for driftless control-affine systems per the Chow–Rashevskii theorem.

Now we will apply this to Langevin dynamics. Assume that $M = I$ for simplicity. Then we can use any of the Lyapunov functions in the family $H^l(q, p) = \left(\frac{1}{2}||p||^2 + U(q)\right)^l$ for l a positive integer, or powers of the Hamiltonian of the system. We allow l to vary so that the condition $|f| \leq H^l$ can hold for many observables f.

To verify that Hörmander's condition holds, 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 $\rho_i \in \mathbb{R}^d$ proportional to e_i, the ith unit vector. Then we see

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

where DY is the Jacobian of Y and d^2U is the hessian of U, so $\{X_1, \ldots, X_d, [X_1, Y], \ldots, [X_d, Y]\}$ spans $\mathbb{R}^{2d}$. Therefore, Langevin dynamics is geometrically ergodic, and as we have previously mentioned, $\mathcal{L}_{\mathrm{LD}}^\dagger \rho_\beta = 0$, so the canonical measure is the measure with respect to which it is ergodic.

4 Splitting Methods

4.1 Theory

Recall that we can decompose Langevin dynamics into A, B, and O parts as in Equation 1. If we would like to update the state of the system for a timestep h, we can perform the A, B, and O parts individually, each for a total timestep of h. We do this by solving the differential equations produced from the Langevin equation by assuming that the other two parts are not present. For example, we find the update for the A part when we assume that the B and O parts are not present. This allows up to define the individual

updates with a timestep of h as

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

where R is a vector of d i.i.d. normal random numbers. Then using these individual updates, we can define a family of splitting methods by performing the updates in various sequences, for example we can produce a scheme that performs B, then O, then A, each with a timestep of h, which we denote as $[\![\mathrm{BOA}]\!]$. We can also perform some of the updates in two halves, each for a timestep of $h/2$, as in the following fashion:

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

We define different splitting schemes in this manner because they produce different practical results, and different methods are better for different situations. For example, consider the 1D harmonic oscillator with spring constant Ω^2, so $U(q) = \Omega^2 q^2/2$. Also assume that there is no friction so $\gamma = 0$. We find the long-time 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	$O(h)$

Table 1: Adapted from Table 7.1 in [1]

where the true covariance $\langle q, p\rangle = 0$. So the scheme $[\![\mathrm{ABOBA}]\!]$ is exact in the variance of the position, but it is off from the true variance of the momentum by $O(h^2)$. The $[\![\mathrm{OABAO}]\!]$, on the other hand, is exact in the variance of momentum but off in the variance of position. The $[\![\mathrm{BAOA}]\!]$ scheme is exact in both of the variance, but the covariance is off by $O(h)$.

The averages we compute $\langle\varphi\rangle_h$ of observables are only approximations of $\langle\varphi\rangle$, so we would like to analyze the accuracy of these approximations. Let $\hat{\rho}(q,p)$ be the stationary distribution produced from iterating a numerical scheme, and let $\hat{\mathcal{L}}^\dagger_{[\![\mathrm{OBA}]\!]}$ by the Kolmogorov operator for this scheme. 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, i.e. $\int_D \hat{\rho}\,\mathrm{d}x = \int_D \rho_\beta\,\mathrm{d}x$. We find

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

so if we know f_k, then we can potentially find $\langle \varphi f_k \rangle$ and the error $\langle \varphi \rangle_h - \langle \varphi \rangle$. Therefore, we would like to compute the correction functions f_k.

Let's assume that $\hat{\mathcal{L}}^\dagger = \mathcal{L}_{\mathrm{LD}}^\dagger + h\mathcal{L}_1^\dagger + h^2\mathcal{L}_2^\dagger + O(h^3)$ for some perturbation operators $\mathcal{L}_k^\dagger$. Then in solving for a stationary solution to the Fokker-Planck equation, we have $\hat{\mathcal{L}}^\dagger \hat{\rho} = 0$, or

$$\left(\mathcal{L}_{\mathrm{LD}}^\dagger + h\mathcal{L}_1^\dagger + h^2\mathcal{L}_2^\dagger + O(h^3)\right)\left(\rho_\beta(1 + hf_1 + h^2 f_2 + h^3 f_3 + O(h^4))\right) = 0$$

from which we obtain the first-order approximation $\mathcal{L}_{\mathrm{LD}}^\dagger(\rho_\beta f_1) + \mathcal{L}_1^\dagger \rho_\beta = 0$ as $\mathcal{L}_{\mathrm{LD}}^\dagger \rho_\beta = 0$. If we know $\mathcal{L}_1^\dagger$, then we can solve this at least numerically for f_1.

We will now apply this to the $[\![\mathrm{OBA}]\!]$ scheme. The following equation holds

$$\exp\left(h\hat{\mathcal{L}}_{[\![\mathrm{OBA}]\!]}^\dagger\right) = \exp\left(h\mathcal{L}_{\mathrm{A}}^\dagger\right)\exp\left(h\mathcal{L}_{\mathrm{B}}^\dagger\right)\exp\left(h\mathcal{L}_{\mathrm{O}}^\dagger\right)$$

where

$$\mathcal{L}_{\mathrm{A}}^\dagger = -p\frac{\partial}{\partial q} \qquad \mathcal{L}_{\mathrm{B}}^\dagger = \nabla U(q)\frac{\partial}{\partial p} \qquad \mathcal{L}_{\mathrm{O}}^\dagger = \gamma(\mathrm{Id} + p\frac{\partial}{\partial p}) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial p^2}$$

Then the Baker–Campbell–Hausdorff (BCH) formula gives

$$\hat{\mathcal{L}}_{[\![\mathrm{OBA}]\!]}^\dagger = \mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{B}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger + \frac{h}{2}\left([\mathcal{L}_{\mathrm{A}}^\dagger, \mathcal{L}_{\mathrm{B}}^\dagger] + [\mathcal{L}_{\mathrm{A}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger] + [\mathcal{L}_{\mathrm{B}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger]\right) + O(h^2)$$

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

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

Therefore, $\hat{\mathcal{L}}^\dagger$ has the requisite deomposition for this scheme to be amenable to the correction function method!

4.2 A Numerical Experiment

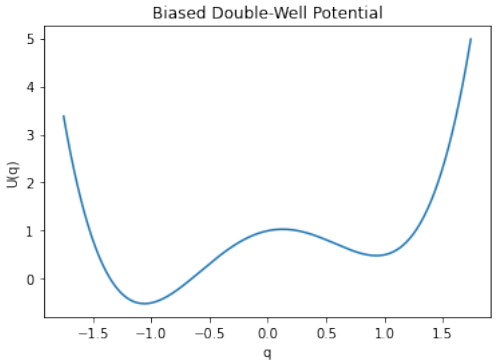

Figure 1: $U(q)$

To illustrate how the correction function method can aid our understanding, consider the biased double-well potential defined as $U(q) = (q^2 - 1)^2 + q/2$ and shown in Figure 1. Let $M = \beta = 1$, and let the observable of interest be $v(q, p) = p^2 - qU'(q) + 2qp$. We can compute $\langle v \rangle = 0$.

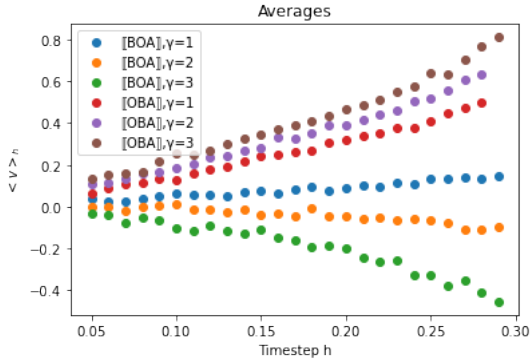

Figure 2: Averages over 3 Trials, each over a total time of 10000 units. This Figure was based on Figure 7.4 in [1], though the data was generated independently.

Numerically simulating Langevin dynamics for this system under the ⟦BOA⟧ and ⟦OBA⟧ schemes yields results as shown in Figure 2. We observe that for $\gamma = 2$, the ⟦BOA⟧ scheme has very low error, and we can explain this rigorously using the correction function method because

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

where $c = \gamma\langle U(q)\rangle$. Therefore, the first order correction is 0 for $\gamma = 2$, and hence we are observing accurate results up to second order! This demonstrates that in practice, the choice of method can depend on the value of γ.

5 Conclusion

Langevin dynamics arises naturally from physical considerations, and is geometrically ergodic. We can find various splitting methods for its SDE, and these splitting methods can be analyzed with correction functions. Lastly, the proofs of these facts use many of the same tools we studied this semester in ESE 559.

In the future, we would like to apply these splitting methods to a biomolecule like how they are applied in [1]. We would also like to theoretically compare the splitting methods presented in this article with other splitting schemes like Stochastic Position Verlet (SPV) and the Brünger-Brooks-Karplus (BBK) integrator. Lastly, we will try to understand how the property of $\mathcal{L}_{\text{LD}}^{\dagger}$ having a compact resolvent ties into the discussion of ergodicity and theoretical properties of Langevin dynamics.

References

[1] Leimkuhler, B., & Matthews, C. (2016). *Molecular Dynamics*. Springer International PU.

[2] 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.

[3] Meyn, S. P., & Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability, 4*(4), 981-1011.

[4] Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association, 90*(430), 558-566.