Error estimates in molecular dynamics

Gabriel STOLTZ

(CERMICS, Ecole des Ponts & MATHERIALS team, Inria Paris)

Research funded by ANR SINEQ and ERC Synergy EMC2

CECAM sandpit research school

Outline

Practical computation of static properties

- Elements of statistical physics
- Ergodic averages using Langevin dynamics

Standard Monte Carlo methods

- Techniques for independent sampling
- Error estimates

A practical introduction to stochastic differential equations

- Brownian motion and diffusion processes
- Discretization of stochastic differential equations
- Langevin-type dynamics

Error estimates for Langevin dynamics

- Numerical discretization
- Types of errors and their scaling

General references (1)

- Computational Statistical Physics
  - D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications* (2002)
  - M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (2010)
  - M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (2017)
  - D. C. Rapaport, *The Art of Molecular Dynamics Simulations* (1995)
  - T. Schlick, *Molecular Modeling and Simulation* (2002)

- Computational Statistics [my personal references... many more out there!]
  - J. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2008
  - W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice* (Chapman & Hall, 1996)

- Machine learning and sampling
  - C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006)
  - K.P. Murphy, *Probabilistic Machine Learning: An Introduction* (MIT Press, 2022)

General references (2)

- Sampling the canonical measure
  - L. Rey-Bellet, Ergodic properties of Markov processes, *Lecture Notes in Mathematics*, **1881** 1–39 (2006)
  - E. Cancès, F. Legoll and G. Stoltz, Theoretical and numerical comparison of some sampling methods, *Math. Model. Numer. Anal.* **41**(2) (2007) 351-390
  - T. Lelièvre, M. Rousset and G. Stoltz, *Free Energy Computations: A Mathematical Perspective* (Imperial College Press, 2010)
  - B. Leimkuhler and C. Matthews, *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods* (Springer, 2015).
  - T. Lelièvre and G. Stoltz, Partial differential equations and stochastic methods in molecular dynamics, *Acta Numerica* **25**, 681-880 (2016)

- Convergence of Markov chains
  - S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability* (Cambridge University Press, 2009)
  - R. Douc, E. Moulines, P. Priouret and P. Soulier, *Markov Chains* (Springer, 2018)

Practical computation of average properties

Microscopic description of physical systems: unknowns

- Microstate of a classical system of N particles:

$$(q,p) = (q_1,\dots,q_N,\ p_1,\dots,p_N) \in \mathcal{E}$$

Positions q (configuration), momenta p (to be thought of as $M\dot{q}$)

- Here, periodic boundary conditions: $\mathcal{E} = \mathcal{D} \times \mathbb{R}^{3N}$ with $\mathcal{D} = (L\mathbb{T})^{3N}$

- Hamiltonian $H(q,p) = E_{\mathrm{kin}}(p) + V(q)$, where the kinetic energy is

$$E_{\mathrm{kin}}(p) = \frac{1}{2}\, p^\top M^{-1} p, \qquad M = \begin{pmatrix} m_1\,\mathrm{Id}_3 & & 0 \\ & \ddots & \\ 0 & & m_N\,\mathrm{Id}_3 \end{pmatrix}$$

All the physics is contained in V

Average properties

- Macrostate of the system described by a probability measure

Equilibrium thermodynamic properties (pressure,...)

$$\mathbb{E}_\mu(\varphi) = \int_{\mathcal{E}} \varphi(q,p)\,\mu(dq\,dp)$$

- Examples of observables:
  - Pressure $\varphi(q,p) = \dfrac{1}{3|\mathcal{D}|}\displaystyle\sum_{i=1}^N \left(\frac{p_i^2}{m_i} - q_i \cdot \nabla_{q_i} V(q)\right)$
  - Kinetic temperature $\varphi(q,p) = \dfrac{1}{3Nk_\mathrm{B}}\displaystyle\sum_{i=1}^N \frac{p_i^2}{m_i}$
- Canonical ensemble = measure on (q,p) (average energy fixed)

$$\mu_{\mathrm{NVT}}(dq\,dp) = Z_{\mathrm{NVT}}^{-1}\,\mathrm{e}^{-\beta H(q,p)}\,dq\,dp, \qquad \beta = \frac{1}{k_\mathrm{B}T}$$

Computing average properties

Main issue

Computation of high-dimensional integrals... Ergodic averages

$$\mathbb{E}_\mu(\varphi) = \lim_{t\to+\infty} \widehat{\varphi}_t, \qquad \widehat{\varphi}_t = \frac{1}{t}\int_0^t \varphi(q_s, p_s)\, ds$$

- One possible choice: Langevin dynamics with friction parameter $\gamma > 0$

= Stochastic perturbation of the Hamiltonian dynamics

$$\begin{cases} dq_t = M^{-1} p_t\, dt \\ dp_t = -\nabla V(q_t)\, dt - \gamma M^{-1} p_t\, dt + \sqrt{\dfrac{2\gamma}{\beta}}\, dW_t \end{cases}$$

Almost-sure convergence of ergodic averages[1]

[1] Kliemann, *Ann. Probab.* **15**(2), 690-707 (1987)

Standard Monte Carlo methods

Standard techniques to sample probability measures (1)

- The basis is the generation of numbers uniformly distributed in $[0, 1]$
- Deterministic sequences which look like they are random...
    - Early methods: linear congruential generators ("chaotic" sequences)

$$x_{n+1} = ax_n + b \mod c, \qquad u_n = \frac{x_n}{c-1}$$

    - Known defects: short periods, point alignments, etc, which can be (partially) patched by cleverly combining several generators

- More recent algorithms: shift registers, such as Mersenne-Twister

$\rightarrow$ defaut choice in *e.g.* Scilab, available in the GNU Scientific Library

- Randomness tests: various flavors

Standard techniques to sample probability measures (2)

- Classical distributions are obtained from the uniform distribution by...
  - inversion of the cumulative function $F(x) = \int_{-\infty}^{x} f(y)\,dy$ (which is an increasing function from $\mathbb{R}$ to $[0,1]$)

$$X = F^{-1}(U) \sim f(x)\,dx$$

Proof: $\mathbb{P}\{a < X \leqslant b\} = \mathbb{P}\{a < F^{-1}(X) \leqslant b\} = \mathbb{P}\{F(a) < U \leqslant F(b)\} = F(b) - F(a) = \int_a^b f(x)\,dx$

Example: exponential law of density $\lambda \mathrm{e}^{-\lambda x} \mathbf{1}_{\{x \geqslant 0\}}$, $F(x) = \mathbf{1}_{\{x \geqslant 0\}}(1 - \mathrm{e}^{-\lambda x})$, so that $X = -\frac{1}{\lambda} \ln U$

  - change of variables: standard Gaussian $G = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$

Proof: $\mathbb{E}(f(X,Y)) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x,y)\,\mathrm{e}^{-(x^2+y^2)/2}\,dx\,dy = \int_0^{+\infty} f\left(\sqrt{r}\cos\theta, \sqrt{r}\sin\theta\right) \frac{1}{2}\mathrm{e}^{-r/2}\,dr\,\frac{d\theta}{2\pi}$

  - using the rejection method

Find a probability density g and a constant $c \geqslant 1$ such that $0 \leqslant f(x) \leqslant c g(x)$. Generate i.i.d. variables $(X^n, U^n) \sim g(x)\,dx \otimes \mathcal{U}[0,1]$, compute $r^n = \frac{f(X^n)}{c g(X^n)}$, and accept X^n if $r^n \geqslant U^n$

Standard techniques to sample probability measures (3)

- The previous methods work only
  - for low-dimensional probability measures
  - when the normalization constants of the probability density are known

- In more complex cases, one needs to resort to trajectory averages

Ergodic methods

$$\frac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(x^n) \xrightarrow[N_{\mathrm{iter}} \to +\infty]{} \int \varphi \, d\mu$$

- **Find methods for which**
  - the convergence is guaranteed? (and in which sense?)
  - error estimates are available? (typically with Central Limit Theorem)

Standard techniques to sample probability measures (4)

- Assume that $x^n \sim \pi$ are idependently and identically distributed (i.i.d.)

Law of Large Numbers for $\varphi \in L^1(\pi)$

$$\widehat{S}_{N_{\mathrm{iter}}} = \frac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(x^n) \xrightarrow[N_{\mathrm{iter}} \to +\infty]{} \mathbb{E}_\pi(\varphi) = \int_{\mathcal{X}} \varphi \, d\pi \quad \text{almost surely}$$

Central Limit Theorem for $\varphi \in L^2(\pi)$

$$\sqrt{N_{\mathrm{iter}}} \left(S_{N_{\mathrm{iter}}} - \int \varphi \, d\pi \right) \xrightarrow[N_{\mathrm{iter}} \to +\infty]{\mathrm{law}} \mathcal{N}(0, \sigma_\varphi^2), \ \ \sigma_\varphi^2 = \int_{\mathcal{X}} \left[\varphi - \mathbb{E}_\pi(\varphi)\right]^2 \, d\pi$$

- Should be thought of as $\widehat{S}_{N_{\mathrm{iter}}} \simeq \mathbb{E}_\pi(\varphi) + \dfrac{\sigma_\varphi}{\sqrt{N_{\mathrm{iter}}}} \mathcal{G}$ with $\mathcal{G} \sim \mathcal{N}(0,1)$

A (practical) introduction to SDEs

Langevin dynamics

- Stochastic perturbation of the Hamiltonian dynamics : friction $\gamma > 0$

$$\begin{cases} dq_t = M^{-1} p_t \, dt \\ dp_t = -\nabla V(q_t) \, dt - \gamma M^{-1} p_t \, dt + \sqrt{\dfrac{2\gamma}{\beta}} \, dW_t \end{cases}$$

- **Motivations**
  - Ergodicity can be proved and is indeed observed in practice
  - Many useful extensions
- **Aims**
  - Understand the meaning of this equation
  - Understand why it samples the canonical ensemble
  - Implement appropriate discretization schemes
  - Estimate the errors (systematic biases vs. statistical uncertainty)

An intuitive view of the Brownian motion (1)

- Independant Gaussian increments whose variance is proportional to time

$$\forall\, 0 < t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n, \qquad W_{t_{i+1}} - W_{t_i} \sim \mathcal{N}(0, t_{i+1} - t_i)$$

where the increments $W_{t_{i+1}} - W_{t_i}$ are independent

- $G \sim \mathcal{N}(m, \sigma^2)$ distributed according to the probability density

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

- The solution of $dq_t = \sigma dW_t$ can be thought of as the limit $\Delta t \to 0$

$$q^{n+1} = q^n + \sigma\sqrt{\Delta t}\, G^n, \qquad G^n \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

where q^n is an approximation of $q_{n\Delta t}$

- Note that $q^n \sim \mathcal{N}(q^0, \sigma^2 n\Delta t)$

- Multidimensional case: $W_t = (W_{1,t}, \dots, W_{d,t})$ where W_i are independent

An intuitive view of the Brownian motion (2)

- Analytical study of the process: law $\psi(t,q)$ of the process at time t

$\rightarrow$ distribution of all possible realizations of q_t for

  - a given initial distribution $\psi(0,q)$, *e.g.* δ_{q^0}
  - and all realizations of the Brownian motion

Averages at time t

$$\mathbb{E}\Big(\varphi(q_t)\Big) = \int_{\mathcal{D}} \varphi(q)\,\psi(t,q)\,dq$$

- Partial differential equation governing the evolution of the law

Fokker-Planck equation

$$\partial_t \psi = \frac{\sigma^2}{2}\Delta\psi$$

Here, simple heat equation $\rightarrow$ "diffusive behavior"

An intuitive view of the Brownian motion (3)

- Proof: Taylor expansion, beware random terms of order $\sqrt{\Delta t}$

$$\varphi\left(q^{n+1}\right) = \varphi\left(q^n + \sigma\sqrt{\Delta t}\, G^n\right)$$
$$= \varphi\left(q^n\right) + \sigma\sqrt{\Delta t} G^n \cdot \nabla\varphi\left(q^n\right) + \frac{\sigma^2 \Delta t}{2}\left(G^n\right)^\top \left(\nabla^2\varphi\left(q^n\right)\right) G^n + \mathrm{O}\left(\Delta t^{3/2}\right)$$

Taking expectations (Gaussian increments G^n independent from the current position q^n)

$$\mathbb{E}\left[\varphi\left(q^{n+1}\right)\right] = \mathbb{E}\left[\varphi\left(q^n\right) + \frac{\sigma^2\Delta t}{2}\Delta\varphi\left(q^n\right)\right] + \mathrm{O}\left(\Delta t^{3/2}\right)$$

Therefore, $\mathbb{E}\left[\dfrac{\varphi\left(q^{n+1}\right) - \varphi\left(q^n\right)}{\Delta t} - \dfrac{\sigma^2}{2}\Delta\varphi\left(q^n\right)\right] \to 0$. On the other hand,

$$\mathbb{E}\left[\frac{\varphi\left(q^{n+1}\right) - \varphi\left(q^n\right)}{\Delta t}\right] \to \partial_t\left(\mathbb{E}\left[\varphi(q_t)\right]\right) = \int_{\mathcal{D}} \varphi(q)\partial_t\psi(t,q)\, dq.$$

This leads to

$$0 = \int_{\mathcal{D}} \varphi(q)\partial_t\psi(t,q)\, dq - \frac{\sigma^2}{2}\int_{\mathcal{D}} \Delta\varphi(q)\,\psi(t,q)\, dq = \int_{\mathcal{D}} \varphi(q)\left(\partial_t\psi(t,q) - \frac{\sigma^2}{2}\Delta\psi(t,q)\right) dq$$

This equality holds for all observables φ.

General SDEs (1)

- State of the system $X \in \mathbb{R}^d$, m-dimensional Brownian motion, diffusion matrix $\sigma \in \mathbb{R}^{d\times m}$

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t$$

to be thought of as the limit as $\Delta t \to 0$ of (X^n approximation of $X_{n\Delta t}$)

$$X^{n+1} = X^n + \Delta t\, b\left(X^n\right) + \sqrt{\Delta t}\,\sigma(X^n)G^n, \qquad G^n \sim \mathcal{N}\left(0, \mathrm{Id}_m\right)$$

- Generator

$$\mathcal{L} = b(x)\cdot\nabla + \frac{1}{2}\sigma\sigma^\top(x) : \nabla^2 = \sum_{i=1}^d b_i(x)\partial_{x_i} + \frac{1}{2}\sum_{i,j=1}^d \left[\sigma\sigma^\top(x)\right]_{i,j}\partial_{x_i}\partial_{x_j}$$

- Proceeding as before, it can be shown that

$$\partial_t\Big(\mathbb{E}\left[\varphi(X_t)\right]\Big) = \int_{\mathcal{X}} \varphi\,\partial_t\psi = \mathbb{E}\Big[\left(\mathcal{L}\varphi\right)\left(X_t\right)\Big] = \int_{\mathcal{X}} \left(\mathcal{L}\varphi\right)\psi$$

General SDEs (2)

Fokker-Planck equation

$$\partial_t \psi = \mathcal{L}^\dagger \psi$$

where $\mathcal{L}^\dagger$ is the adjoint of $\mathcal{L}$

$$\int_{\mathcal{X}} (\mathcal{L}\varphi)(x)\, B(x)\, dx = \int_{\mathcal{X}} \varphi(x) \left(\mathcal{L}^\dagger B\right)(x)\, dx$$

- Invariant measures are stationary solutions of the Fokker-Planck equation

Invariant probability measure $\psi_\infty(x)\, dx$

$$\mathcal{L}^\dagger \psi_\infty = 0, \qquad \int_{\mathcal{X}} \psi_\infty(x)\, dx = 1, \qquad \psi_\infty \geqslant 0$$

- When $\mathcal{L}$ is elliptic (*i.e.* $\sigma\sigma^\top$ has full rank: the noise is sufficiently rich), the process can be shown to be irreducible = accessibility property

$$P_t(x, \mathcal{S}) = \mathbb{P}(X_t \in \mathcal{S} \mid X_0 = x) > 0$$

General SDEs (3)

- Sufficient conditions for ergodicity
  - irreducibility
  - existence of an invariant probability measure $\psi_\infty(x)\,dx$

Then the invariant measure is unique and

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(X_t)\,dt = \int_{\mathcal{X}} \varphi(x)\,\psi_\infty(x)\,dx \qquad \text{a.s.}$$

- Rate of convergence given by Central Limit Theorem: $\widetilde{\varphi} = \varphi - \displaystyle\int \varphi\,\psi_\infty$

$$\sqrt{T}\left(\frac{1}{T}\int_0^T \varphi(X_t)\,dt - \int_{\mathcal{X}} \varphi\,\psi_\infty\right) \xrightarrow[T\to+\infty]{\text{law}} \mathcal{N}(0,\sigma_\varphi^2)$$

with $\sigma_\varphi^2 = 2\,\mathbb{E}\left[\displaystyle\int_0^{+\infty} \widetilde{\varphi}(X_t)\widetilde{\varphi}(X_0)dt\right]$ (proof: later, discrete time setting)

SDEs: numerics (1)

- Numerical discretization: various schemes (Markov chains in all cases)

- Example: Euler–Maruyama

$$X^{n+1} = X^n + \Delta t \, b(X^n) + \sqrt{\Delta t} \, \sigma(X^n) \, G^n, \qquad G^n \sim \mathcal{N}(0, \mathrm{Id}_d)$$

- Standard notions of error: fixed integration time $T < +\infty$
  - Strong error $\displaystyle \sup_{0 \leqslant n \leqslant T/\Delta t} \mathbb{E}|X^n - X_{n\Delta t}| \leqslant C \Delta t^a$
  - Weak error: $\displaystyle \sup_{0 \leqslant n \leqslant T/\Delta t} \Big| \mathbb{E}\left[\varphi\left(X^n\right)\right] - \mathbb{E}\left[\varphi\left(X_{n\Delta t}\right)\right] \Big| \leqslant C \Delta t^a$ (for any φ)
  - "mean error" *vs.* "error of the mean"

- Example: for Euler–Maruyama, weak order 1, strong order $1/2$ (1 when σ constant)

SDEs: numerics (2)

- Trajectorial averages: estimator $\widehat{\Phi}_{N_{\mathrm{iter}}} = \dfrac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(X^n)$

- Numerical scheme ergodic for the probability measure $\psi_{\infty,\Delta t}$

- Two types of errors to compute averages w.r.t. invariant measure
  - Statistical error, quantified using a Central Limit Theorem

$$\widehat{\Phi}_{N_{\mathrm{iter}}} = \int_{\mathcal{X}} \varphi\, \psi_{\infty,\Delta t} + \frac{\sigma_{\Delta t,\varphi}}{\sqrt{N_{\mathrm{iter}}}} \mathscr{G}_{N_{\mathrm{iter}}}, \qquad \mathscr{G}_{N_{\mathrm{iter}}} \sim \mathcal{N}(0,1)$$

  - Systematic errors
    - perfect sampling bias, related to the finiteness of Δt

$$\left| \int_{\mathcal{X}} \varphi\, \psi_{\infty,\Delta t} - \int_{\mathcal{X}} \varphi\, \psi_{\infty} \right| \leqslant C_\varphi\, \Delta t^a$$

    - finite sampling bias, related to the finiteness of N_{iter}

Expression of the asymptotic variance: correlations matter!

$$\sigma^2_{\Delta t,\varphi} = \mathrm{Var}(\varphi) + 2\sum_{n=1}^{+\infty} \mathbb{E}\Big(\widetilde{\varphi}(X^n)\widetilde{\varphi}(X^0)\Big), \qquad \widetilde{\varphi} = \varphi - \int \varphi\, \psi_{\infty,\Delta t}$$

where $\mathrm{Var}(\varphi) = \displaystyle\int_{\mathcal{X}} \widetilde{\varphi}^2 \psi_{\infty,\Delta t} = \int_{\mathcal{X}} \varphi^2 \psi_{\infty,\Delta t} - \left(\int_{\mathcal{X}} \varphi \psi_{\infty,\Delta t}\right)^2$

- Note also that $\sigma^2_{\Delta t,\varphi} \approx \dfrac{2}{\Delta t}\mathbb{E}\left[\displaystyle\int_0^{+\infty} \widetilde{\varphi}(X_t)\widetilde{\varphi}(X_0)\, dt\right]$

Proof: by the stationary property $\mathbb{E}\left[\varphi(q^n)\varphi(q^m)\right] = \mathbb{E}\left[\varphi(q^{n-m})\varphi(q^0)\right]$

$$N_{\mathrm{iter}}\mathbb{E}\left(\widehat{\Phi}^2_{N_{\mathrm{iter}}}\right) = \frac{1}{N_{\mathrm{iter}}}\sum_{n=0}^{N_{\mathrm{iter}}-1}\mathbb{E}\left[\varphi(q^n)^2\right] + \frac{2}{N_{\mathrm{iter}}}\sum_{0\leqslant m<n\leqslant N_{\mathrm{iter}}-1}\mathbb{E}\left[\varphi(q^n)\varphi(q^m)\right]$$

$$= \mathbb{E}\left(\varphi^2\right) + 2\sum_{1\leqslant n\leqslant N_{\mathrm{iter}}-1}\left(1-\frac{n}{N_{\mathrm{iter}}}\right)\mathbb{E}\left[\varphi(q^n)\varphi(q^0)\right]$$

Overdamped Langevin dynamics

- SDE on the configurational part only (momenta trivial to sample)

$$dq_t = -\nabla V(q_t)\,dt + \sqrt{\frac{2}{\beta}}dW_t$$

- Invariance of the canonical measure $\nu(dq) = \psi_0(q)\,dq$

$$\psi_0(q) = Z^{-1}\,\mathrm{e}^{-\beta V(q)}, \qquad Z = \int_{\mathcal{D}} \mathrm{e}^{-\beta V}$$

- Generator $\mathcal{L} = -\nabla V(q)\cdot\nabla_q + \dfrac{1}{\beta}\Delta_q$
  - invariance of ψ_0: adjoint $\mathcal{L}^\dagger\varphi = \mathrm{div}_q\left((\nabla V)\varphi + \dfrac{1}{\beta}\nabla_q\varphi\right)$
  - elliptic generator hence irreducibility and ergodicity
- Discretization $q^{n+1} = q^n - \Delta t\,\nabla V(q^n) + \sqrt{\dfrac{2\Delta t}{\beta}}\,G^n$ (+ Metropolization)

Langevin dynamics (1)

- Stochastic perturbation of the Hamiltonian dynamics

$$\begin{cases} dq_t = M^{-1} p_t \, dt \\ dp_t = -\nabla V(q_t)\, dt - \gamma M^{-1} p_t \, dt + \sigma \, dW_t \end{cases}$$

- γ,σ may be matrices, and may depend on q

- Generator $\mathcal{L} = \mathcal{L}_{\rm ham} + \mathcal{L}_{\rm thm}$

$$\mathcal{L}_{\rm ham} = p^\top M^{-1} \nabla_q - \nabla V(q)^\top \nabla_p = \sum_{i=1}^{dN} \frac{p_i}{m_i} \partial_{q_i} - \partial_{q_i} V(q) \partial_{p_i}$$

$$\mathcal{L}_{\rm thm} = -p^\top M^{-1} \gamma^\top \nabla_p + \frac{1}{2} \left(\sigma \sigma^\top \right) : \nabla_p^2 \qquad \left(= \frac{\sigma^2}{2} \Delta_p \text{ for scalar } \sigma \right)$$

- Irreducibility can be proved (control argument)

Langevin dynamics (2)

- Invariance of the canonical measure to conclude to ergodicity?

Fluctuation/dissipation relation

$$\sigma\sigma^\top = \frac{2}{\beta}\gamma \qquad \text{implies} \qquad \mathcal{L}^\dagger\left(\mathrm{e}^{-\beta H}\right) = 0$$

- Proof for scalar γ, σ: a simple computation shows that

$$\mathcal{L}_{\mathrm{ham}}^\dagger = -\mathcal{L}_{\mathrm{ham}}, \qquad \mathcal{L}_{\mathrm{ham}} H = 0$$

- Overdamped Langevin analogy $\mathcal{L}_{\mathrm{thm}} = \gamma\left(-p^\top M^{-1}\nabla_p + \frac{1}{\beta}\Delta_p\right)$

$\rightarrow$ Replace q by p and $\nabla V(q)$ by $M^{-1}p$

$$\mathcal{L}_{\mathrm{thm}}^\dagger\left[\exp\left(-\beta\frac{p^\top M^{-1}p}{2}\right)\right] = 0$$

- Conclusion: $\mathcal{L}_{\mathrm{ham}}^\dagger$ and $\mathcal{L}_{\mathrm{thm}}^\dagger$ both preserve $\mathrm{e}^{-\beta H(q,p)}\,dq\,dp$

Langevin dynamics (3)

- **Asymptotic variance** $\sigma_\varphi^2 = \lim_{t\to+\infty} t\,\mathrm{Var}_\mu(\widehat{\varphi}_t)$: with $\Pi\varphi = \varphi - \int_{\mathcal{E}} \varphi\, d\mu$,

$$\sigma_\varphi^2 = \lim_{t\to+\infty} \int_0^t \left(1 - \frac{s}{t}\right) \mathbb{E}_\mu \left[\Pi\varphi(q_t,p_t)\Pi\varphi(q_0,p_0)\right] ds$$
$$= 2\int_0^{+\infty} \int_{\mathcal{E}} (\mathrm{e}^{s\mathcal{L}}\Pi\varphi)\Pi\varphi\, d\mu\, ds = 2\int_{\mathcal{E}} (-\mathcal{L}^{-1}\Pi\varphi)\Pi\varphi\, d\mu$$

Well-defined provided $-\mathcal{L}\Phi = \Pi\varphi$ has a solution in $L_0^2(\mu) = \Pi L^2(\mu)$

A Central Limit Theorem holds in this case[2]: $\boxed{\widehat{\varphi}_t - \mathbb{E}_\mu(\varphi) \simeq \frac{\sigma_\varphi}{\sqrt{t}}\mathcal{G}}$

- Sufficient condition: integrability of the semigroup, *e.g.*

$$\left\|\mathrm{e}^{t\mathcal{L}}\right\|_{\mathcal{B}(L_0^2(\mu))} \leqslant C\mathrm{e}^{-\lambda t}$$

so that $-\mathcal{L}^{-1} = \int_0^{+\infty} \mathrm{e}^{s\mathcal{L}}\, ds$

[2] R. N. Bhattacharya, *Z. Wahrsch. Verw. Gebiete* (1982)

Langevin dynamics (4)

Prove exponential convergence of the semigroup $\mathrm{e}^{t\mathcal{L}}$ on $E \subset L_0^2(\mu)$

- Lyapunov techniques[3] $L^\infty_{\mathscr{K}}(\mathcal{E}) = \left\{ \varphi \,\text{measurable}, \, \sup \left| \dfrac{\varphi}{\mathscr{K}} \right| < +\infty \right\}$
- standard hypocoercive[4] setup $H^1(\mu)$
- $L^2(\mu)$ after hypoelliptic regularization[5] from $H^1(\mu)$
- direct transfer from $H^1(\mu)$ to $L^2(\mu)$ by spectral argument[6]
- directly[7] $L^2(\mu)$ (recently[8] Poincaré using $\partial_t - \mathcal{L}_{\mathrm{ham}}$)
- coupling arguments[9]
- direct estimates on the resolvent using Schur complements[10]

Rate of convergence $\min\left(\gamma, \gamma^{-1}\right)$ so variance $\sim \max\left(\gamma, \gamma^{-1}\right)$

---

[3] Wu ('01); Mattingly/Stuart/Higham ('02); Rey-Bellet ('06); Hairer/Mattingly ('11)

[4] Villani (2009) and before Talay (2002), Eckmann/Hairer (2003), Hérau/Nier (2004),...

[5] Hérau, *J. Funct. Anal.* (2007)

[6] Deligiannidis/Paulin/Doucet, *Ann. Appl. Probab.* (2020)

[7] Hérau (2006), Dolbeaut/Mouhot/Schmeiser (2009, 2015)

[8] Armstrong/Mourrat (2019), Cao/Lu/Wang (2019), Brigatti (2021), Brigati/Stoltz (2023)

[9] Eberle/Guillin/Zimmer, *Ann. Probab.* (2019)

[10] Bernard/Fathi/Levitt/Stoltz, *Annales Henri Lebesgue* (2022)

Hamiltonian and overdamped limits

- As $\gamma \to 0$, the Hamiltonian dynamics is recovered
$$\frac{d}{dt}\mathbb{E}\left[H(q_t,p_t)\right] = -\gamma\left(\mathbb{E}\left[p_t^\top M^{-2} p_t\right] - \frac{1}{\beta}\mathrm{Tr}(M^{-1})\right) dt$$

Time $\sim \gamma^{-1}$ to change energy levels in this limit[11]

- Overdamped limit $\gamma \to +\infty$ with $M = \mathrm{Id}$: rescaling of time γt
$$\begin{aligned} q_{\gamma t} - q_0 &= -\frac{1}{\gamma}\int_0^{\gamma t} \nabla V(q_s)\, ds + \sqrt{\frac{2}{\gamma\beta}} W_{\gamma t} - \frac{1}{\gamma}\left(p_{\gamma t} - p_0\right) \\ &= -\int_0^t \nabla V(q_{\gamma s})\, ds + \sqrt{2\beta^{-1}} B_t - \frac{1}{\gamma}\left(p_{\gamma t} - p_0\right) \end{aligned}$$

which converges to the solution of $dQ_t = -\nabla V(Q_t)\, dt + \sqrt{2\beta^{-1}}\, dB_t$

- Alternatively, $\mathrm{e}^{\gamma t(\mathcal{L}_{\mathrm{ham}} + \gamma \mathcal{L}_{\mathrm{FD}})} \approx \mathrm{e}^{t\mathcal{L}_{\mathrm{ovd}}}$ with $\mathcal{L}_{\mathrm{ovd}} = -\nabla V^\top \nabla_q + \beta^{-1}\Delta_q$

- In both cases, slow convergence, with rate scaling as $\min\left(\gamma, \gamma^{-1}\right)$

[11] Hairer and Pavliotis, *J. Stat. Phys.*, **131**(1), 175-202 (2008)

Error estimates on the computation of average properties

Numerical integration of the Langevin dynamics (1)

- Splitting strategy: Hamiltonian part + fluctuation/dissipation

$$\left\{\begin{aligned} dq_t &= \ M^{-1}\, p_t\, dt \\ dp_t &= -\nabla V(q_t)\, dt \end{aligned}\right. \qquad \oplus \qquad \left\{\begin{aligned} dq_t &= 0 \\ dp_t &= -\gamma M^{-1} p_t\, dt + \sqrt{\frac{2\gamma}{\beta}}\, dW_t \end{aligned}\right.$$

- Hamiltonian part integrated using a Verlet scheme

- Analytical integration of the fluctuation/dissipation part

$$d\left(\mathrm{e}^{\gamma M^{-1} t} p_t\right) = \mathrm{e}^{\gamma M^{-1} t}\left(dp_t + \gamma M^{-1} p_t\, dt\right) = \sqrt{\frac{2\gamma}{\beta}} \mathrm{e}^{\gamma M^{-1} t}\, dW_t$$

so that

$$p_t = \mathrm{e}^{-\gamma M^{-1} t}\, p_0 + \sqrt{\frac{2\gamma}{\beta}} \int_0^t \mathrm{e}^{-\gamma M^{-1}(t-s)}\, dW_s$$

It can be shown that $\displaystyle\int_0^t f(s)\, dW_s \sim \mathcal{N}\left(0, \int_0^t f(s)^2 ds\right)$

Numerical integration of the Langevin dynamics (2)

- Trotter splitting (define $\alpha_{\Delta t} = \mathrm{e}^{-\gamma M^{-1}\Delta t}$, choose $\gamma M^{-1}\Delta t \sim 0.01 - 1$)

$$
\begin{cases}
p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla V(q^n), \\
q^{n+1} = q^n + \Delta t\, M^{-1} p^{n+1/2}, \\
\widetilde{p}^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}), \\
p^{n+1} = \alpha_{\Delta t}\widetilde{p}^{n+1} + \sqrt{\dfrac{1-\alpha_{2\Delta t}}{\beta} M}\, G^n,
\end{cases}
$$

Error estimate on the invariant measure $\mu_{\Delta t}$ of the numerical scheme

There exist a function f such that, for any smooth observable φ,

$$
\int_{\mathcal{E}} \varphi\, d\mu_{\Delta t} = \int_{\mathcal{E}} \varphi\, d\mu + \Delta t^2 \int_{\mathcal{E}} \varphi\, f\, d\mu + \mathrm{O}(\Delta t^3)
$$

- Strang splitting: same accuracy

Practical computation of average properties

- Numerical scheme = Markov chain characterized by evolution operator

$$P_{\Delta t}\varphi(q,p) = \mathbb{E}\Big(\varphi\left(q^{n+1},p^{n+1}\right)\,\Big|\,(q^n,p^n)=(q,p)\Big)$$

- Discretization of the Langevin dynamics: splitting strategy

$$A = M^{-1}p\cdot\nabla_q, \qquad B = -\nabla V(q)\cdot\nabla_p, \qquad O = -M^{-1}p\cdot\nabla_p + \frac{1}{\beta}\Delta_p$$

- First order splitting schemes: $P_{\Delta t}^{ZYX} = \mathrm{e}^{\Delta t Z}\mathrm{e}^{\Delta t Y}\mathrm{e}^{\Delta t X} \simeq \mathrm{e}^{\Delta t \mathcal{L}}$
- Example: $P_{\Delta t}^{B,A,O}$ corresponds to (with $\alpha_{\Delta t} = \exp(-\gamma M^{-1}\Delta t)$)

$$\begin{cases} \widetilde{p}^{n+1} = p^n - \Delta t\,\nabla V(q^n), \\ q^{n+1} = q^n + \Delta t\, M^{-1}\widetilde{p}^{n+1}, \\ p^{n+1} = \alpha_{\Delta t}\widetilde{p}^{n+1} + \sqrt{\dfrac{1-\alpha_{\Delta t}^2}{\beta}M}\,G^n, \end{cases} \tag{1}$$

where G^n are i.i.d. standard Gaussian random variables

Practical computation of average properties (2)

- Second order splitting $P_{\Delta t}^{ZYXYZ} = \mathrm{e}^{\Delta t Z/2} \mathrm{e}^{\Delta t Y/2} \mathrm{e}^{\Delta t X} \mathrm{e}^{\Delta t Y/2} \mathrm{e}^{\Delta t Z/2}$
- Example: $P_{\Delta t}^{O,B,A,B,O}$ (Verlet in the middle)

$$
\begin{cases}
\widetilde{p}^{n+1/2} = \alpha_{\Delta t/2} p^n + \sqrt{\dfrac{1-\alpha_{\Delta t}}{\beta} M}\, G^n, \\[2mm]
p^{n+1/2} = \widetilde{p}^{n+1/2} - \dfrac{\Delta t}{2} \nabla V(q^n), \\[2mm]
q^{n+1} = q^n + \Delta t\, M^{-1} p^{n+1/2}, \\[2mm]
\widetilde{p}^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2} \nabla V(q^{n+1}), \\[2mm]
p^{n+1} = \alpha_{\Delta t/2} \widetilde{p}^{n+1} + \sqrt{\dfrac{1-\alpha_{\Delta t}}{\beta} M}\, G^{n+1/2},
\end{cases}
$$

- Other category: Geometric Langevin algorithms, *e.g.* $P_{\Delta t}^{O,A,B,A}$
- **Current recommendation: BAOAB scheme**

B. Leimkuhler and Ch. Matthews, *Appl. Math. Res. Express* (2013)
N. Bou-Rabee and H. Owhadi, *SIAM J. Numer. Anal.* (2010)

Types of errors

Estimators of $\mathbb{E}_\mu(\varphi)$

$$\widehat{\varphi}_t = \frac{1}{t}\int_0^t \varphi(q_s,p_s)\,ds, \qquad \widehat{\varphi}_{\Delta t}^{N_{\rm iter}} = \frac{1}{N_{\rm iter}}\sum_{n=1}^{N_{\rm iter}} \varphi(q^n,p^n)$$

Statistical error (variance of the estimator) : $\mathrm{O}\left(\frac{1}{\sqrt{N_{\rm iter}\Delta t}}\right)$

- dictated by the central limit theorem for continuous dynamics
- discrete dynamics: asymptotic variance coincides at order Δt^a

Bias (expectation of the estimator)

- finite time integration $\to$ bias $\mathrm{O}\left(\frac{1}{N_{\rm iter}\Delta t}\right)$
- discretization of the dynamics $\to$ bias $\mathrm{O}(\Delta t^a)$

Finite time integration bias

Bias $\mathrm{O}(1/t)$, typically smaller than statistical error $\mathrm{O}(1/\sqrt{t})$

$$\left|\mathbb{E}\left(\widehat{\varphi}_t\right) - \mathbb{E}_\mu(\varphi)\right| \leqslant \frac{K}{t}$$

Key equality for the proofs: introduce $-\mathcal{L}\Phi = \Pi\varphi$ and write

$$\begin{aligned}\widehat{\varphi}_t - \mathbb{E}_\mu(\varphi) &= \frac{1}{t}\int_0^t \Pi\varphi(q_s,p_s)\,ds \\ &= \frac{\Phi(q_0,p_0)-\Phi(q_t,p_t)}{t} + \sqrt{\frac{2\gamma}{\beta}}\,\frac{1}{t}\int_0^t \nabla_p\Phi(q_s,p_s)^\top dW_s\end{aligned}$$

with Ito calculus $d\Phi(q_s,p_s) = \mathcal{L}\Phi(q_s,p_s) + \sqrt{2\gamma\beta^{-1}}\nabla_p\Phi(q_s,p_s)^\top dW_s$

Also allows to prove CLT: martingale part dominant, with variance

$$\frac{2\gamma}{\beta t^2}\int_0^t \mathbb{E}\left[\left|\nabla_p\Phi(q_s,p_s)\right|^2\right] ds \sim \frac{2\gamma}{\beta t}\int_{\mathcal{E}} \left|\nabla_p\Phi\right|^2\, d\mu = \frac{2\gamma}{\beta t}\int_{\mathcal{E}} \Phi(-\mathcal{L}\Phi)\,d\mu$$

Timestep discretization bias

The ergodicity of numerical schemes can be proved ($\mathcal{D}$ bounded):

$$\frac{1}{N_{\text{iter}}} \sum_{n=1}^{N_{\text{iter}}} \varphi(q^n, p^n) \xrightarrow[N_{\text{iter}} \to +\infty]{} \int \varphi(q,p)\, d\mu_{\gamma,\Delta t}(q,p)$$

Systematic error estimates: α order of the splitting scheme

$$\int_{\mathcal{E}} \varphi(q,p)\, \mu_{\gamma,\Delta t}(dq\, dp) = \int_{\mathcal{E}} \varphi(q,p)\, \mu(dq\, dp) + \Delta t^{\alpha} \int_{\mathcal{E}} \varphi(q,p) f_{\alpha,\gamma}(q,p)\, \mu(dq\, dp) + \mathrm{O}(\Delta t^{\alpha+1})$$

Correction function $f_{\alpha,\gamma}$ solution of an appropriate Poisson equation

$$\mathcal{L}^* f_{\alpha,\gamma} = g_\gamma$$

where g_γ depends on the numerical scheme (adjoints taken on $L^2(\mu)$)

Proof for the first-order scheme $P_{\Delta t}^{O,B,A}$ (1)

- By definition of the invariant measure, $\displaystyle\int_{\mathcal{E}} P_{\Delta t}\phi \, d\mu_{\gamma,\Delta t} = \int_{\mathcal{E}} \phi \, d\mu_{\gamma,\Delta t}$, so

$$\int_{\mathcal{E}} \left[\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t} \right) \phi \right] d\mu_{\gamma,\Delta t} = 0$$

- In view of the BCH formula $\mathrm{e}^{\Delta t A_3} \mathrm{e}^{\Delta t A_2} \mathrm{e}^{\Delta t A_1} = \mathrm{e}^{\Delta t \mathcal{A}}$ with

$$\mathcal{A} = A_1 + A_2 + A_3 + \frac{\Delta t}{2} \Big([A_3, A_1 + A_2] + [A_2, A_1] \Big) + \dots,$$

it holds $P_{\Delta t}^{O,B,A} = \mathrm{Id} + \Delta t \mathcal{L} + \dfrac{\Delta t^2}{2} \left(\mathcal{L}^2 + S_1 \right) + \Delta t^3 R_{1,\Delta t}$ with

$$S_1 = [C, A + B] + [B, A], \qquad R_{1,\Delta t} = \frac{1}{2} \int_0^1 (1-\theta)^2 \mathcal{R}_{\theta \Delta t} \, d\theta,$$

Proof for the first-order scheme $P_{\Delta t}^{O,B,A}$ (2)

- The correction function $f_{1,\gamma}$ is chosen so that

$$\int_{\mathcal{E}} \left[\left(\frac{\mathrm{Id} - P_{\Delta t}^{O,B,A}}{\Delta t} \right) \phi \right] (1 + \Delta t f_{1,\gamma}) \, d\mu = \mathrm{O}(\Delta t^2)$$

This requirement can be rewritten as

$$0 = \int_{\mathcal{E}} \left(\frac{1}{2} S_1 \phi + (\mathcal{L}\phi) f_{1,\gamma} \right) d\mu = \int_{\mathcal{E}} \varphi \left[\frac{1}{2} S_1^* \mathbf{1} + \mathcal{L}^* f_{1,\gamma} \right] d\mu,$$

which suggests to choose $\mathcal{L}^* f_{1,\gamma} = -\dfrac{1}{2} S_1^* \mathbf{1}$ (well posed equation)

- Technical work to replace $\left(\dfrac{\mathrm{Id} - P_{\Delta t}^{O,B,A}}{\Delta t} \right) \phi$ by φ

Conclusion and perspectives

Need for variance reduction methods

Dominant error = statistical error

$$\frac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(q^n, p^n) \approx \int_{\mathcal{E}} \varphi \, d\mu + \frac{\sigma_\varphi}{\sqrt{N_{\mathrm{iter}} \Delta t}} \mathcal{G} + \dots$$

Variance reduction: play on σ_φ^2 (cannot play on the scaling)

- stratification (Umbrella sampling)
- importance sampling (replace V by $V + \widetilde{V}$ and reweight trajectory)
- control variates (replace φ by $\varphi + \mathcal{L}\Phi$)