An introduction to high dimensional sampling

Gabriel STOLTZ

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

MAC-MIGS tutorial "High Dimensional Sampling and Applications"

Outline

- **Examples of high-dimensional probability measures**
  - Statistical physics
  - Bayesian inference

Practical computation of average properties

- Ergodic averages using Langevin dynamics
- Sources of errors: bias and variance

Timestep bias for the computation of average properties

- Discretization of Langevin dynamics
- A priori estimates on the invariant measure

General references (1)

Computational Statistical Physics

- D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications* (Academic Press, 2002)
- M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford, 2010)
- M.P. Allen and D.J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 1987)

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)
- C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer, 2004)

Machine learning and sampling

- D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, 2012)
- C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006)
- K.P. Murphy, *Probabilistic Machine Learning: An Introduction* (MIT Press, 2022)

General references (2)

Mathematical works on sampling (Gibbs) measures

- 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)

Examples of high dimensional probability measures

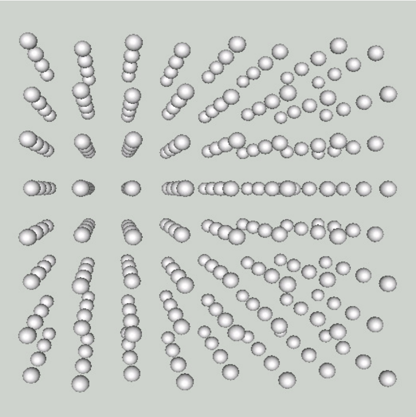
Statistical physics (1)

- **Aims of computational statistical physics**
  - numerical microscope
  - computation of average properties, static or dynamic

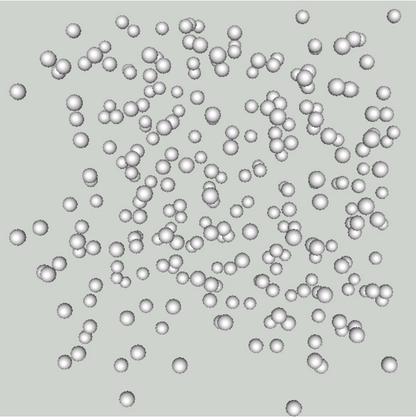
- **Orders of magnitude**
  - distances $\sim 1\ \mathring{A} = 10^{-10}$ m
  - energy per particle $\sim k_{\mathrm{B}}T \sim 4 \times 10^{-21}$ J at room temperature
  - atomic masses $\sim 10^{-26}$ kg
  - time $\sim 10^{-15}$ s
  - number of particles $\sim \mathcal{N}_A = 6.02 \times 10^{23}$

- **"Standard" simulations**
  - 10^6 particles ["world records": around 10^9 particles]
  - integration time: (fraction of) ns ["world records": (fraction of) μs]

What is the melting temperature of argon?

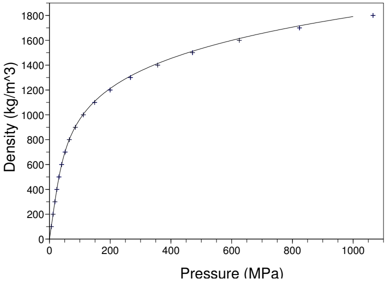


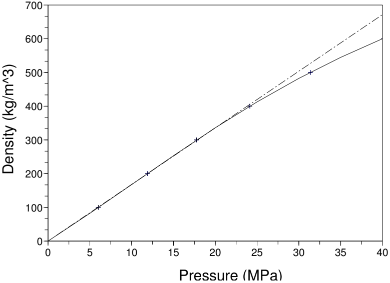
(a) Solid argon (low temperature)



(b) Liquid argon (high temperature)

"Given the structure and the laws of interaction of the particles, what are the macroscopic properties of the matter composed of these particles?"

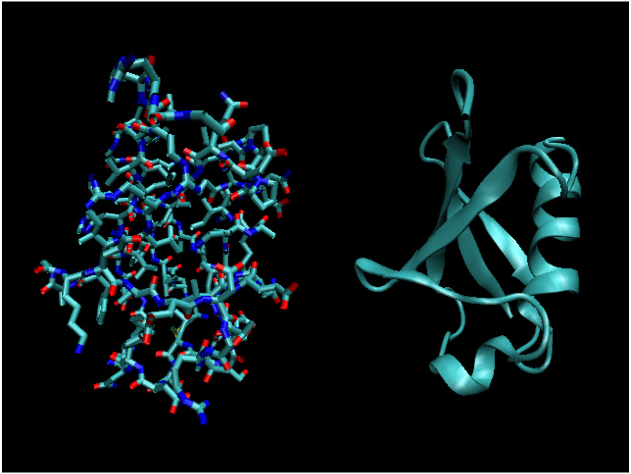




Equation of state (pressure/density diagram) for argon at $T = 300$ K

Statistical physics (4)

What is the structure of the protein? What are its typical conformations, and what are the transition pathways from one conformation to another?



Statistical physics (5)

• 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 $M\dot{q}$)

• In the simplest cases, $\mathcal{E} = \mathcal{D} \times \mathbb{R}^{3N}$ with $\mathcal{D} = \mathbb{R}^{3N}$ or $\mathbb{T}^{3N}$

• More complicated situations can be considered: molecular constraints defining submanifolds of the phase space

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

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

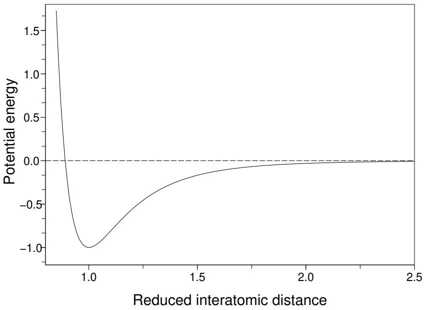
- All the physics is contained in V
  - ideally derived from quantum mechanical computations
  - in practice, empirical potentials for large scale calculations

- An example: Lennard-Jones pair interactions to describe noble gases

$$V(q_1,\dots,q_N) = \sum_{1\leqslant i<j\leqslant N} v(|q_j - q_i|)$$

$$v(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

$$\text{Argon:} \quad \begin{cases} \sigma = 3.405\times 10^{-10}\ \text{m} \\ \varepsilon/k_{\text{B}} = 119.8\ \ \text{K} \end{cases}$$



Statistical physics (7)

- Macrostate of the system described by a probability measure

Equilibrium thermodynamic properties (pressure,...)

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

- Choice of thermodynamic ensemble
  - least biased measure compatible with the observed macroscopic data
  - Volume, energy, number of particles, ... fixed exactly or in average
  - Equivalence of ensembles (as $N \to +\infty$)

- Canonical ensemble = measure on (q,p), average energy fixed H

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

with $\beta = \dfrac{1}{k_{\mathrm{B}}T}$ the Lagrange multiplier of the constraint $\displaystyle\int_{\mathcal{E}} H\,\rho\,dq\,dp = E_0$

Bayesian inference (1)

- Data set $\{x_i\}_{i=1,\dots,N_{\rm data}}$
- Elementary likelihood $P(x|q)$, with q parameters of probability measure
- A priori distribution of the parameters $p_{\rm prior}$ (usually not so informative)

Aim

Find the values of the parameters q describing correctly the data: sample

$$\nu(q) \propto p_{\rm prior}(q) \prod_{i=1}^{N_{\rm data}} P(x_i|q)$$

- Example of Gaussian mixture model

Bayesian inference (2)

- Elementary likelihood approximated by mixture of K Gaussians

$$P(x\,|\,\theta) = \sum_{k=1}^{K} a_k \sqrt{\frac{\lambda_k}{2\pi}} \exp\left(-\frac{\lambda_k}{2}(x-\mu_k)^2\right)$$

- Parameters $\theta = (a_1, \dots, a_{K-1}, \mu_1, \dots, \mu_K, \lambda_1, \dots, \lambda_K)$ with

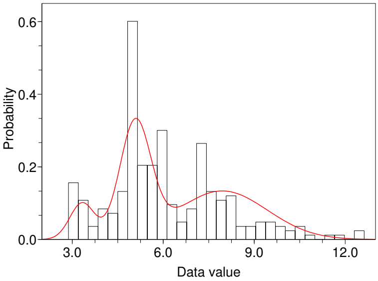
$$\mu_k \in \mathbb{R}, \quad \lambda_k \geqslant 0, \quad 0 \leqslant a_k \leqslant 1, \quad a_1 + \dots + a_K = 1$$

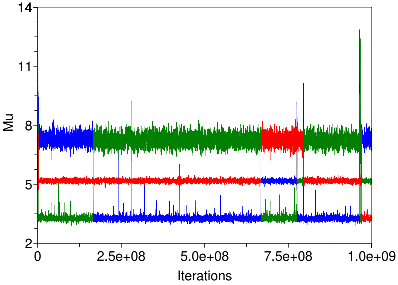
- Prior distribution: Random beta model: additional variable
  - uniform distribution of the weights a_k
  - $\mu_k \sim \mathcal{N}\left(M, R^2/4\right)$ with $M =$ mean of data, $R = \max y_i - \min y_i$
  - $\lambda_k \sim \Gamma(\alpha, \beta)$ with $\beta \sim \Gamma(g, h)$, $g = 0.2$ and $h = 100g/\alpha R^2$

Aim

Find the values of the parameters (namely θ, and possibly K as well) describing correctly the data

[RG97] S. Richardson and P. J. Green. *J. Roy. Stat. Soc. B*, 1997.
[JHS05] A. Jasra, C. Holmes and D. Stephens, Statist. Science, 2005



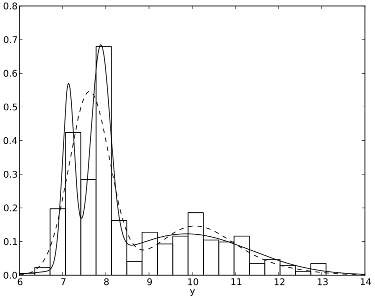


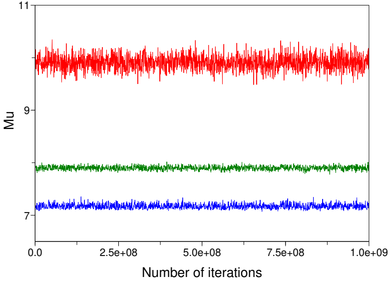
Left: Lengths of snappers ($N_{\mathrm{data}} = 256$), and a possible fit for $K = 3$ using the last configuration from the trajectory plotted in the right picture.

Right: Typical sampling trajectory, Metropolis/Gaussian random walk with $(\sigma_q, \sigma_\mu, \sigma_v, \sigma_\beta) = (0.0005, 0.025, 0.05, 0.005)$.

[IS88] A. J. Izenman and C. J. Sommer, *J. Am. Stat. Assoc.*, 1988.
[BMY97] K. Basford *et al.*, *J. Appl. Stat.*, 1997

Bayesian inference (4)



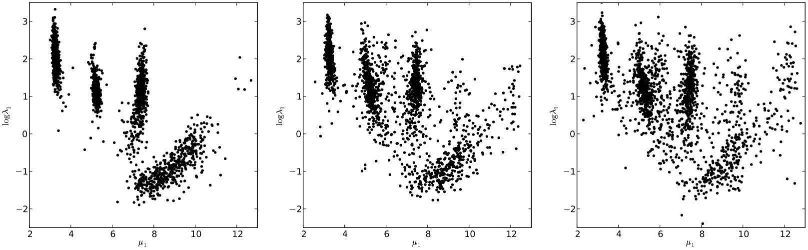


Left: Thickness of Mexican stamps ("Hidalgo stamp data", $N_{\rm data} = 485$), and two possible fits for $K = 3$ ("genuine multimodality", solid line: dominant mode).

Right: Typical sampling trajectory

[TSM86] D. Titterington *et al.*, *Statistical Analysis of Finite Mixture Distributions*, 1986.
[FS06] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, 2006.

Bayesian inference (5)



Scatter plot of the marginal distribution of $(\mu_1, \log \lambda_1)$ for the Fish data, for various values of $K = 4, 5, 6$

Computing average properties

In all cases: Target measure $\nu(dq) = Z_\nu^{-1} \mathrm{e}^{-\beta V(q)} \, dq$
Extended measure $\mu(dq\,dp) = \nu(dq)\kappa(dp)$ with marginal ν

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

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

- Other choices include Metropolis-like schemes

Practical computation of average properties

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

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

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 (1D case to simplify)

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

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

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

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

where X^n is an approximation of $X_{n\Delta t}$

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

- 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 $\psi(t,x)$ of the process at time t
$\rightarrow$ distribution of all possible realizations of X_t for

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

Averages at time t

$$\mathbb{E}\Big(\varphi(X_t)\Big) = \int_{\mathcal{X}} \varphi(x)\,\psi(t,x)\,dx$$

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"

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

$$\begin{aligned}\varphi\left(X^{n+1}\right) &= \varphi\left(X^n + \sigma\sqrt{\Delta t}\, G^n\right) \\ &= \varphi\left(X^n\right) + \sigma\sqrt{\Delta t} G^n \cdot \nabla\varphi\left(X^n\right) + \frac{\sigma^2 \Delta t}{2}\left(G^n\right)^T \left(\nabla^2 \varphi\left(X^n\right)\right) G^n + \mathrm{O}\left(\Delta t^{3/2}\right)\end{aligned}$$

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

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

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

$$\mathbb{E}\left[\frac{\varphi\left(X^{n+1}\right) - \varphi\left(X^n\right)}{\Delta t}\right] \to \partial_t\left(\mathbb{E}\left[\varphi(X_t)\right]\right) = \int_{\mathcal{X}} \varphi(x)\partial_t\psi(t,x)\, dx$$

This leads to

$$0 = \int_{\mathcal{X}} \varphi(x)\partial_t\psi(t,x)\, dx - \frac{\sigma^2}{2}\int_{\mathcal{X}} \Delta\varphi(x)\, \psi(t,x)\, dx = \int_{\mathcal{X}} \varphi(x)\left(\partial_t\psi(t,x) - \frac{\sigma^2}{2}\Delta\psi(t,x)\right) dx$$

This equality holds for all observables φ

General SDEs (1)

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

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

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

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

Proceeding as before, it can be shown that

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

General SDEs (2)

Fokker-Planck equation

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

where $\mathcal{L}^\dagger$ is the flat L^2 adjoint of $\mathcal{L}$

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

Invariant measures: stationary solutions of the Fokker-Planck equation

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

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

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

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

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: $\widetilde{\varphi} = \varphi - \displaystyle\int \varphi\,\psi_\infty$

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

with $\sigma_\varphi^2 = 2\,\mathbb{E}\left[\displaystyle\int_0^{+\infty} \widetilde{\varphi}(X_t)\widetilde{\varphi}(X_0)dt\right] \longrightarrow$ error of order $\dfrac{\sigma_\varphi}{\sqrt{T}}$

SDEs: numerics (1)

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

Standard notions of error: fixed integration time $T < +\infty$

- Strong error $\displaystyle\sup_{0 \leqslant n \leqslant T/\Delta t} \mathbb{E}|X^n - X_{n\Delta t}| \leqslant C\Delta t^p$
- Weak error: $\displaystyle\sup_{0 \leqslant n \leqslant T/\Delta t} \Big| \mathbb{E}\left[\varphi\left(X^n\right)\right] - \mathbb{E}\left[\varphi\left(X_{n\Delta t}\right)\right] \Big| \leqslant C_\varphi \Delta t^p$
- "mean error" *vs.* "error of the mean"

Example: Euler-Maruyama scheme

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

weak order 1, strong order $1/2$ (1 when σ constant)

SDEs: numerics (2)

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

$$\Phi_{N_{\rm iter}} = \frac{1}{N_{\rm iter}} \sum_{n=1}^{N_{\rm iter}} \varphi(X^n)$$

Two types of errors to compute averages w.r.t. invariant measure

- Statistical error, quantified using a Central Limit Theorem

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

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

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

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

SDEs: numerics (3)

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

Key point: *The statistical error coincides at dominant order in* Δt *with the one of the continous process on the same timescale*

$$\Delta t \sigma^2_{\Delta t,\varphi} \sim 2\mathbb{E}\left[\int_0^{+\infty} \widetilde{\varphi}(X_t)\widetilde{\varphi}(X_0)\, dt\right] = \sigma^2_{\varphi}$$

Estimation: block averaging, approximation of integrated autocorrelation

B. Leimkuhler, C. Matthews and G. Stoltz, *IMA J. Numer. Anal.* (2016)
T. Lelièvre and G. Stoltz, *Acta Numerica* (2016)

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 $\nu(dq) = \psi_0(q)\, dq$

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

Generator $\mathcal{L} = -\nabla V(q) \cdot \nabla_q + \frac{1}{\beta}\Delta_q$

- invariance of ψ_0: adjoint $\mathcal{L}^\dagger \varphi = \mathrm{div}_q \left((\nabla V)\varphi + \frac{1}{\beta}\nabla_q \varphi \right)$
- elliptic generator hence irreducibility and ergodicity

Discretization $q^{n+1} = q^n - \Delta t\, \nabla V(q^n) + \sqrt{\frac{2\Delta t}{\beta}}\, G^n$ (+ 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 $\mathcal{L} = \mathcal{L}_{\mathrm{ham}} + \mathcal{L}_{\mathrm{thm}}$

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

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

Irreducibility can be proved (control argument)

Langevin dynamics (2)

Invariance of the canonical measure to conclude to ergodicity?

Fluctuation/dissipation relation

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

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

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

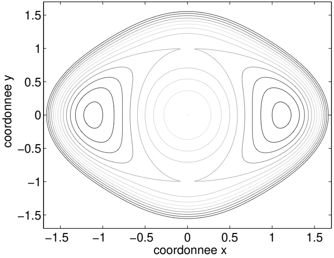
Overdamped Langevin analogy $\mathcal{L}_{\mathrm{thm}} = \gamma\left(-p^T M^{-1}\nabla_p + \frac{1}{\beta}\Delta_p\right)$

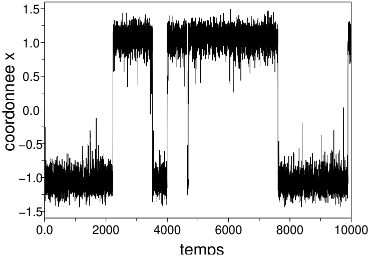
$\to$ Replace q by p and $\nabla V(q)$ by $M^{-1}p$

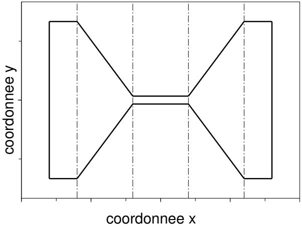
$$\mathcal{L}_{\mathrm{thm}}^\dagger\left[\exp\left(-\beta\frac{p^T M^{-1} p}{2}\right)\right] = 0$$

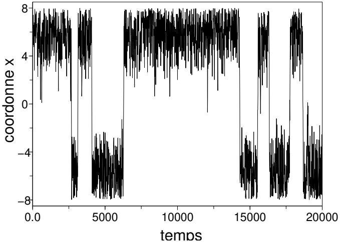
Conclusion: $\mathcal{L}_{\mathrm{ham}}^\dagger$ and $\mathcal{L}_{\mathrm{thm}}^\dagger$ both cancel $\mathrm{e}^{-\beta H(q,p)}\,dq\,dp$

Metastability: large statistical error...









CLT discussed tomorrow. Need for variance reduction techniques!

Timestep bias for the computation of average properties

Practical computation of average properties (1)

- 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

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

- 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: $P_{\Delta t}^{B,A,\gamma C}$ 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: $P_{\Delta t}^{\gamma C,B,A,B,\gamma C}$ (Verlet in the middle)

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

- Other category: Geometric Langevin algorithms, *e.g.* $P_{\Delta t}^{\gamma C,A,B,A}$

Error estimates on the computation of average properties

Ergodicity of numerical schemes — Durmus/Enfroy/Moulines/Stoltz (2021)

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

Statistical errors *vs.* systematic errors (bias)

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 Poisson equation (scheme specific)

Talay/Tubaro (1990) for the general strategy, Leimkuhler/Matthews (2013), Abdulle/Vilmart/Zygalakis (2014), Leimkuhler/Matthews/Stoltz (2016)

Proof for the first-order scheme $P_{\Delta t}^{\gamma C,B,A}$ (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 $P_{\Delta t}^{\gamma C,B,A} = \mathrm{Id} + \Delta t \mathcal{L} + \dfrac{\Delta t^2}{2}\left(\mathcal{L}^2 + S_1 \right) + \Delta t^3 R_{1,\Delta t}$ 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 $P_{\Delta t}^{\gamma C,B,A}$ (2)

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

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

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)

- Replace ϕ by $\left(\dfrac{\mathrm{Id} - P_{\Delta t}^{\gamma C,B,A}}{\Delta t} \right)^{-1} \varphi$? No control on the derivatives...

- Rely on the "nice" properties of the continuous dynamics, *i.e.* functional estimates[1] on $\mathcal{L}^{-1}$ to use pseudo-inverses

$$Q_{1,\Delta t} = -\mathcal{L}^{-1} + \frac{\Delta t}{2} (\mathrm{Id} + \mathcal{L}^{-1} S_1 \mathcal{L}^{-1})$$

[1] D. Talay, Stoch. Proc. Appl. (2002); M. Kopec (2015)

Summary and next steps

Where we are...

What we did...

- Examples of high-dimensional probability measures
- Practical computation of average properties (ergodic averages with Langevin dynamics)
- General discussion of errors (statistical vs. systematic)
- Timestep bias for the computation of average properties

What we will do tomorrow...

hypocoercive techniques for the convergence of semigroups generated by degenerate elliptic operators

(motivation = Fokker–Planck equation for Langevin dynamics, to make sense of the asymptotic variance)