Convergence to equilibrium under variable curvature bounds.

Max Fathi
joint work with Patrick Cattiaux (Toulouse) and Arnaud Guillin (Clermont-Ferrand)

24 octobre 2020

Drift-diffusion processes

Consider the noisy gradient descent equation

$$dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t.$$

Markov process, with generator $L = \Delta - \nabla V \cdot \nabla$. Assume $\int e^{-V} dx = 1$. Then $\mu(dx) = e^{-V(x)}dx$ is the unique invariant probability measure.

Question : how fast does the law of X_t converge to μ ?

Motivations from statistical physics and numerical simulations.

Spectral gap

The operator $-L$ is symmetric in $L^2(\mu)$, and long-time behavior of the process is related to its spectrum.

$$\int (Lf) g d\mu = -\int \nabla f \cdot \nabla g d\mu$$

so a spectral gap for $-L$ is equivalent to a Poincaré inequality

$$\mathrm{Var}_\mu(f) \leq C_P \int |\nabla f|^2 d\mu,$$

in which case $1/C_P$ is a lower bound on the positive eigenvalues of $-L$. As soon as a Poincaré inequality holds, if $\nu_t = f_t \mu$ is the law of X_t we have

$$||f_t - 1||_{L^2(\mu)} \leq e^{-t/C_P} ||f_0 - 1||_{L^2(\mu)}.$$

Connection with concentration of measure

Poincaré inequalities implies sub-exponential concentration properties : for any 1-lipschitz function

$$\mu(f \geq \mu(f) + r) \leq 3 \exp\left(-\frac{r}{\sqrt{C_P}}\right).$$

Converse not true in general, but true if V is convex (E. Milman 2008).

Tensorization property : $C_P(\mu^{\otimes n}) = C_P(\mu)$, so behaves well in high dimension.

Theorem (Lichnerowicz '54, Brascamp-Lieb '79, Bakry-Emery '85, Caffarelli '01)

If Hess $V \geq \alpha\,\mathrm{Id}$ *with* $\alpha > 0$, *then* $C_P \leq \alpha^{-1}$.

Sharp bound in general, since it gives the best constant for Gaussian distributions.

Dimension-free behavior, crucial for many applications.

Also works in the geometric setting, assuming Ricci curvature lower bounds.

One way of proving this bound is via the Bakry-Emery gradient estimate $|\nabla P_t f| \leq e^{-\alpha t} P_t |\nabla f|$.
We have

$$Var_\mu(f) = 2\int_0^\infty \int |\nabla P_t f|^2 d\mu dt \leq 2\int_0^\infty \int e^{-2\alpha t} P_t |\nabla f|^2 d\mu dt$$

$$= \alpha^{-1} \int |\nabla f|^2 d\mu.$$

More generally, to prove a Poincaré inequality, it is enough to show that for any 1-lipschitz f and $t \geq 0$

$$|\nabla P_t f| \leq ce^{-t/C_P}.$$

Another consequence of the gradient bound is the convergence to equilibrium in W_1 distance.

$$W_1(\mu,\nu) = \sup_{f\,1-lip} \int f d\mu - \int f d\nu = \inf_{\pi} \int |x-y| d\pi.$$

Then for any initial data ν_0, we have

$$W_1(\nu_t,\mu) \leq \sup_{f\,1-lip} \int P_t f d\nu_0 - \int P_t f d\mu \leq e^{-\alpha t} W_1(\nu_0,\mu).$$

Often a better metric than L^2 convergence for high-dimensional applications.

Question : how to weaken the requirement of a uniform lower bound on Hess V ?

Starting point : general representation of $\nabla P_t f$. If Hess $V(x) \geq \alpha(x)$ for some function α that is bounded from below, then

$$|\nabla P_t f(x)| \leq \mathbb{E}\left[|\nabla f(X_t)| \exp\left(-\int_0^t \alpha(X_s)ds\right)\right].$$

Particular case of an identity proved by Braun, Habermann and Sturm in the geometric setting. This kind of formula goes back to Bismut's work on Malliavin calculus (1984).

Recovers the Bakry-Emery gradient bound when f is lipschitz and α bounded from below.

Sketch of proof

We couple two processes with initial conditions x and y using the *same* Brownian motion. Using the lower bound

$$\langle \nabla V(z) - \nabla V(z'), z - z' \rangle \geq \int_0^1 \alpha(\lambda z + (1-\lambda)z')|z - z'|^2 d\lambda$$

we can get

$$e^{\int_0^t 2 \int_0^1 \alpha(\lambda X_s^x + (1-\lambda)X_s^y)|z-z'|^2 d\lambda ds}|X_t^x - X_t^y|^2 \leq |x - y|^2.$$

We can then get the gradient bound by testing out against functions, and letting y go to x.

If we only consider 1-lipschitz functions, we have

$$|\nabla P_t f(x)| \leq \mathbb{E}_x \left[\exp\left(-\int_0^t \alpha(X_s) ds \right) \right].$$

So to prove either convergence to equilibrium in W_1 distance or some Poincaré inequality, we aim to get a bound of the form

$$\mathbb{E}_x \left[\exp\left(-\int_0^t \alpha(X_s) ds \right) \right] \leq Ce^{-ct}.$$

By the ergodic theorem for Markov processes, we expect $\int_0^t \alpha(X_s) ds \approx t \int \alpha d\mu$. So can try to get exponential convergence with rate the average curvature.

Two motivations :

- In situations where the curvature is negative in a small region, but positive in most of the space, we expect exponential convergence to equilibrium to still occur. Average curvature may be a good tool in such situations.
- Even when $\alpha(x) \geq \alpha_0 > 0$ for all x, we have $C_P = \alpha_0^{-1}$ iff α is actually constant (Cheng & Zhou '17, Gigli, Ketterer, Kuwada & Ohta '18). So any positive variable curvature bound should lead to a strict improvement of the Bakry-Emery bound. Veysseire (2011) : can use harmonic mean of curvature.

Unfortunately, positive average curvature cannot be a sufficient condition. In dimension one, we have

$$\int V'' e^{-V} dx = \int (V')^2 e^{-V} dx > 0$$

and yet there are potentials V for which convergence to equilibrium may be slower than exponential. So some other condition is necessary.

The gap is that to quantitatively replace $\int_0^t \alpha(X_s) ds$ by $t \int \alpha d\mu$, we need a rate of convergence in the ergodic theorem, at least for the observable α.

Transport-information inequalities

Theorem (Guillin, Léonard, Wu & Yao 2009)

Given a distance d on $\mathbb{R}^d$, the following are equivalent :

- *The measure μ satisfies a transport-information inequality*

$$W_{1,d}(\rho\mu,\mu)^2 \leq C_I \int |\nabla \log \rho|^2 d\mu$$

  w.r.t. the distance d ;

- *For any test function f that is 1-lipschitz with respect to d and with $\int f d\mu = 0$, we have*

$$\mathbb{P}_{\rho\mu}\left[\int_0^t f(X_s)ds \geq rt\right] \leq ||\rho||_{L^2(\mu)} \exp(-tr^2/C_I).$$

  where $\rho\mu$ is the law of the initial datal X_0.

By Markov's inequality, the second condition implies bounds on the probability that $\int_0^t f(X_s)ds$ deviates from the limit predicted by the ergodic theorem.

Two sub-cases of interest here :

- If d is the Euclidean distance, the associated transport-information inequality is weaker than the logarithmic Sobolev inequality, and stronger than Gaussian concentration.
- If d is the discrete distance, the admissible test functions are the bounded functions, and the functional inequality is actually weaker than the Poincaré inequality itself, with $C_I = 4C_P$.

An application : self-improvement in the Bakry-Emery bound

Theorem

Assume that $\text{Hess}\, V(x) \geq \alpha(x) \geq \alpha_0 > 0$. *Then* $C_P \leq (\alpha_0 + t)^{-1}$ *with*

$$t := \left(\mu(\alpha) - \alpha_0 + \alpha_0^{-1}\,\mathrm{Osc}(\alpha)^2\right)\left(1 - \sqrt{1 - \frac{(\mu(\alpha)-\alpha_0)^2}{(\mu(\alpha) - \alpha_0 + \alpha_0^{-1}\,\mathrm{Osc}(\alpha)^2)^2}}\right)$$

Always a strict improvement for measures with full support and no Gaussian factor.

Sketch of proof

- Use the a priori bound $C_P \leq \alpha_0^{-1}$ to bound $\mathbb{P}\left(\int_0^t \alpha(X_s)ds > t(r+\mu(\alpha))\right)$. For lower deviation bounds can use $\alpha \geq \alpha_0$.

- Deduce an a posteriori bound on $\mathbb{E}\left[\exp\left(\int -\alpha(X_s)ds\right)\right]$, and optimize in r.

- Use the gradient estimate to deduce an improved bound on C_P.

Several other ways of implementing this strategy (using log-Sobolev inequalities, median of the curvature bound, other concentration inequalities...).

Can also be used to get convergence to equilibrium in W_p distance with $p > 1$, but with weaker rates, and replacing the curvature lower bound by a smaller function κ such that

$$\langle \nabla V(x) - \nabla V(y), x - y \rangle \geq (\kappa(x) + \kappa(y))|x - y|^2.$$

Can for example be used to get W_p rates with $p > 2$ if a log-Sobolev inequality holds.

Thanks for your attention !