Advanced computational methods-Lecture 4

1 Basic concepts of SDEs: the generator, semi-groups

Let μ_t be the law of $X(t)$, which is a measure in $\mathbb{R}^d$. Then we have

$$\mathbb{E}f(X(t)) = \langle \mu_t, f \rangle = \int_{\mathbb{R}^d} f \, d\mu_t.$$

For smooth bounded function $f(x)$, define

$$u(x,t) = \mathbb{E}_x f(X_t).$$

By Itô's formula, u satisfies

$$\partial_t u(x,t) = \mathbb{E}_x \mathcal{L} f(X_t),$$

where $\mathcal{L}$ is the generator of the process

$$\mathcal{L} := b \cdot \nabla + \frac{1}{2}\Lambda_{ij}\partial_{ij},$$

where we used Einstein summation convention (i.e. $\Lambda_{ij}\partial_{ij} \equiv \sum_{i,j=1}^d \Lambda_{ij}\partial_{x_i x_j}$) and

$$\Lambda = \sigma\sigma^T.$$

This is a special case of Dynkin's formula. The density of the law of $X(t)$ starting x, denoted by $p(t,y;x)$, is called the Green's function. When Λ is positive definite, $p(t,x,y)$ is a smooth function for $t > 0$. Dynkin's equation implies that $p(t,y;x)$ satisfies the **forward Kolmogorov equation, or Fokker-Planck equation** for $t > 0$:

$$\partial_t p = -\nabla_y \cdot (b(y)p) + \frac{1}{2}\partial_{y_i y_j}(\Lambda_{ij}(y)p) := \mathcal{L}_y^* p,$$

where the subindex y means that the derivatives are taken on y variable. By the well-posedness of the SDE, we have

$$\int_{\mathbb{R}^d} p(t,y;x)\, dy = 1, \ \forall x \in \mathbb{R}^d, t > 0.$$

Clearly, for general starting probability measure μ_0, the law of $X(t)$ also satisfies the Fokker-Planck equation in the distributional sense:

$$\frac{d}{dt}\langle \mu_t, f\rangle = \langle \mu_t, \mathcal{L}f\rangle.$$

Moreover, let $v : (x,t) \mapsto v(x,t)$ solve the **backward Kolmogorov equation**

$$\partial_t v = \mathcal{L}v = b \cdot \nabla v + \frac{1}{2}\Lambda_{ij}\partial_{ij} v$$

with initial condition $v(x,0) = f(x)$. Let $X(t)$ be the process satisfying the SDE with initial condition $X(0) = x$. We check that $Y_s = v(X(s), t-s)$ is a martingale and therefore

$$v(x,t) = Y_0 = \mathbb{E}Y_t = \mathbb{E}v(X(t),0) = \mathbb{E}f(X(t)) = u(x,t).$$

This means that u solves the backward Kolmogorov equation. Combining with Dynkin's formula, we can infer that the Green's function satisfies $\mathcal{L}_y^* p(t,x,y) = \mathcal{L}_x p(t,x,y)$, or

$$\begin{aligned} &-\nabla_y \cdot (b(y)p(t,x,y)) + \frac{1}{2}\partial_{y_i y_j}(\Lambda_{ij}(y)p(t,x,y)) = \\ &\qquad\qquad b(x)\cdot\nabla_x p(t,x,y) + \frac{1}{2}\Lambda_{ij}(x)\partial_{x_i x_j} p(t,x,y). \end{aligned} \quad (1)$$

We now define the semigroup generated by $\mathcal{L}$:

$$S_t = e^{t\mathcal{L}},$$

which means

$$\frac{d}{dt}S_t = \mathcal{L}S_t.$$

In other words, let u_0 be a suitable function, then $u(t) = e^{t\mathcal{L}}u_0$ is the solution of the backward Kolmogorov equation at t with initial data u_0.

Similarly, we define

$$S_t^* = e^{t\mathcal{L}^*}.$$

Hence, for a given initial density p_0, then

$$p_t = e^{t\mathcal{L}^*}p_0$$

is the probability density at t.

Using the above representation we have the following observation

Proposition 1. *$e^{t\mathcal{L}^*}$ is nonnegativity preserving, integral preserving and it is L^1 non-expansive ($\|S_t^*\|_{L^1\to L^1} \le 1$). Meanwhile, $e^{t\mathcal{L}}$ is L^∞ non-expansive and nonnegativity preserving.*

The second claim follows from the fact $(L^1)' = L^\infty$. Also can be seen by the representation

$$u(x,t) = \int p(t,y;x)u(y,0)\,dy.$$

In fact, if we look at the PDEs:

$$\partial_t u = \mathcal{L}u = b\cdot\nabla u + \Lambda : \nabla^2 u.$$

On the right hand side there is no u term. Such equations are known to have maximal principle and thus L^∞ non-expansive.

The Fokker-Planck equation is in conservative form

$$\partial_t p = \mathcal{L}^* p.$$

If we expand the derivatives, it has constant term, so it does not have maximal principle.

Remark 1. *Using the semigroup, the relation* (1) *can be understood easily in a formal way. Consider the 1D case:*

$$\begin{aligned}\mathcal{L}_y^* p(y,t;x) = \mathcal{L}_y^* e^{t\mathcal{L}_y^*}\delta(y-x) = e^{t\mathcal{L}_y^*}L_y^*(\delta(y-x)) &= e^{t\mathcal{L}_y^*}[-\partial_y(b(y)\delta(y-x)) + \frac{1}{2}\partial_{yy}(\sigma^2(y)\delta(y-x))]\\ &= e^{t\mathcal{L}_y^*}[-\partial_y(b(x)\delta(y-x)) + \frac{1}{2}\partial_{yy}(\sigma^2(x)\delta(y-x))]\\ &= e^{t\mathcal{L}_y^*}[b(x)\partial_x\delta(y-x) + \sigma^2(x)\frac{1}{2}\partial_{xx}\delta(y-x)] = \mathcal{L}_x p(y,t;x)\end{aligned}$$

2 Dynkin and Feynman-Kac

Above, we have seen a special case of Dynkin's formula. In fact, for stopping time τ with $\mathbb{P}(\tau < \infty) = 1$, the Dynkin's formula still holds

$$\mathbb{E}_x g(X_\tau) = g(x) + \mathbb{E}_x \int_0^\tau \mathcal{L}g(X_s)ds.$$

Here, a stopping time $\tau : \Omega \to [0,\infty)$ is a random variable such that

$$\{\tau \le t\} \in \mathcal{F}_t, \forall t \ge 0.$$

intuitively, this means whether we stop the process at time t or not can be known by the the information given up to t. Typical stopping time is the hitting time of a given domain.

Consider the ellitpic PDE with Dirichlet boundary condition

$$\mathcal{L}u = f, \quad u(x \in \partial D) = \varphi(x)$$

Then, Dynkin's formula tells us that

$$\mathbb{E}_x u(X_\tau) = u(x) + \mathbb{E}_x \int_0^\tau \mathcal{L}u(X_s) ds.$$

If we pick τ to be the hitting time, then $u(X_\tau) = \varphi(X_\tau)$. The second term on the right hand side is $\mathbb{E}_x \int_0^\tau f(X_s) ds$. Hence, overall, we have

$$u(x) = \mathbb{E}_x \varphi(X_\tau) - \mathbb{E}_x \int_0^\tau f(X_s) ds.$$

A particularly interesting case is when $f = -1$ and $\varphi = 0$. This means

$$u(x) = \mathbb{E}_x \tau,$$

solves the PDE

$$\mathcal{L}u = -1, \quad u(x \in \partial D) = 0$$

If the test function depends on time, we should have the follwoing formula

$$\mathbb{E}_x g(\tau, X_\tau) = g(0, x) + \mathbb{E}_x \int_0^\tau (\partial_t g + \mathcal{L}g)(s, X_s)\, ds.$$

The **Feymann-Kac** formula is a generalization of the backward equation. Let

$$v(t, x) = \mathbb{E}_x \left[\exp(-\int_0^t q(X_s)\, ds) \varphi(X_t) \right].$$

then

$$\partial_t v = \mathcal{L}v - qv.$$

Note that in this expectation, X is the normal diffusion with probability conserved. One can also introduce another process $\tilde{X}$ with birth and death. Then, the above expectation could be written as $v(t, x) = \mathbb{E}_x \varphi(\tilde{X})$. Then, q is the rate that the probability of $\tilde{X}$ can be created or be killed.

3 Convergence to equilibrium

Assumption 1. *Suppose b and σ are smooth. The function b satisfies*

$$b(x) \cdot x \leq -r|x|^2 \tag{2}$$

when $|x| > R$ for some R. Also, σ satisfies $\|\sigma\|_\infty < \infty$ and $\Lambda = \sigma\sigma^T \geq S_1 I > 0$.

The process has certain recurrent properties so that the SDE has a unique stationary distribution π [1, sect. 4.4-4.7]. Moreover, π has a density with respect to Lebesgue measure [1, Lemma 4.16]. If $\sigma\sigma^T$ is positive definite everywhere, we have $\pi(y) > 0$.

Note that stationary distribution means that the law of $X(t)$ converges in some sense to a distribution. Or, in other words, the process $X(t)$ converges in distribution to some random variable X_∞. However, $X(t)$ never converges pointwise (there is Brownian motion always, so $X(t, \omega)$ cannot tend to some value). π is also called the invariant measure.

Hence, one desired to show the convergence to the invariant measure. Some ways to prove

- Regarding the convergence of $u(\cdot, t)$ to $\langle \pi, f \rangle$ or μ_t to π using coupling argument for SDEs. In particular, we have the V-uniform geometric ergodicity for $u(\cdot, t) \to \langle \pi, f \rangle$ ([2, 3]) or geometric convergence of $\mu_t \to \pi$ in Wasserstein space ([4, 5]).

- Prove the geometric convergence of $u(\cdot, t)$ to $\langle \pi, f \rangle$ in $L^p(\pi)$ spaces using spectral gap and Perron-Frobenius type theorems (see [2, Chap. 20]; [6, 7, 8] for example).

- PDE techniques. Use the Fokker-Planck equation and some functional inequalities (Poincaré inequality, or log Sobolev inequality etc). These functional inequalities will imply spectral gaps of the semigroups.

Here, I will mention a method using the PDE technique to study a special case:

$$dX = -\nabla V(X)dt + \sqrt{2}dW_s,$$

where V is strongly convex. Some of the calculation here can be found in [9].

In this case, using the Fokker-Planck equation, we can verify that

$$\pi(x) = C \exp(-V(x)),$$

where C is some normalization constant.

Define

$$q(x,t) := \frac{p(x,t)}{\pi(x)} \geq 0,$$

Note that Λ_{ij} is symmetric and

$$-\nabla \cdot (b\pi) + \frac{1}{2}\partial_{ij}(\Lambda_{ij}\pi) = 0, \tag{3}$$

we have

$$\partial_t q = \Big(\frac{1}{\pi}\nabla \cdot (\Lambda\pi) - b\Big) \cdot \nabla q + \frac{1}{2}\Lambda_{ij}\partial_{ij}q. \tag{4}$$

If the detailed balance condition

$$b = \frac{1}{2\pi}\nabla \cdot (\Lambda\pi) \tag{5}$$

holds (for example, $\Lambda = 2DI$ and $b = -\nabla V$), then we have the useful identity

$$\mathcal{L}^*(f\pi) = \pi\mathcal{L}f + f\mathcal{L}^*\pi = \pi\mathcal{L}f. \tag{6}$$

Then (4) can be rewritten as

$$\partial_t q = b \cdot \nabla q + \frac{1}{2}\Lambda_{ij}\partial_{ij}q, \tag{7}$$

which is the backward equation.

Remark 2. *Some people may have doubts why we call* (5) *the detailed balance condition. We will come back to this when we talk about MCMC.*

Now, we take

$$\varphi(q) = q\log q - q + 1,$$

and have

$$\frac{\partial}{\partial t}\varphi(q) = \varphi'(q)\mathcal{L}(q) = \mathcal{L}\varphi(q) - D\varphi''(q)|\nabla q|^2.$$

Hence, we have

$$\frac{d}{dt}\int \varphi(q)\pi\, dx = -D\int \varphi''(q)|\nabla q|^2\pi\, dx$$

We define

$$H(\rho|e^{-V}) = \int \varphi(q)\pi\, dx = \int q\log q\pi\, dx = \int \rho\log\frac{\rho}{Ce^{-V}}\, dx = \int \rho\log\frac{\rho}{e^{-V}}\, dx - C_1$$

This is known as the relative entropy

$$\varphi''(q)|\nabla q|^2\pi = \frac{\pi}{q}|\nabla q|^2 = \rho|\nabla(\log q)|^2 = \rho|\nabla(\log\frac{\rho}{e^{-V}})|^2.$$

We define

$$I(\rho|e^{-V}) = \int \frac{|\nabla q|^2}{q}\pi, dx = \int \rho|\nabla(\log\frac{\rho}{e^{-V}})|^2\,dx$$

to be the relative Fisher information.

Hence,

$$\frac{d}{dt}H(\rho|e^{-V}) = -I(\rho|e^{-V}).$$

Our goal is then to show

$$I(\rho|e^{-V}) \geq 2\lambda H(\rho|e^{-V}).$$

If a measure e^{-V} satisfies this inequality, we say it satisfies a logarithmic Sobolev inequality with constant λ.

Remark 3. *The original Log Sobolev inequality is for Gaussian measure. Let* $\gamma = (2\pi)^{-d/2}\exp(-|x|^2/2)$*. Then,*

$$\int u^2\log u^2\,d\gamma - (\int u^2 d\gamma)\log(\int u^2\,d\gamma) \leq 2\int |\nabla u|^2\,d\gamma.$$

This means we have the embedding $H^1(d\gamma) \subset L^2\log L^2(d\gamma)$*. This Sobolev inequality is then equivalent to*

$$H(\rho|\gamma) \leq \frac{1}{2}I(\rho|\gamma)$$

Here, we have two facts.

The first is the Holley-Stroock perturbation lemma (the paper title is "Logarithmic Sobolev inequalities and stochastic Ising models"):

Lemma 1. *If* $V = V_0 + v$ *where* e^{-V_0} *satisfies the log Sobolev inequality with constant* λ *and* $osc(v) := \sup v - \inf v < \infty$*, then* e^{-V} *satisfies the log Sobolev inequality with constant* $\lambda e^{-osc(v)}$*.*

The second is a more fundamental result by Bakry and Emery in 1985.

Lemma 2. *If* $\nabla^2 V \geq \lambda I_d$*, then* e^{-V} *satisfies the Log Sobolev inequality with constant* λ*.*

Since we aim to be simple and consider strongly convex V, so I will prove the second lemma only.

Proof. We first of all assume $\rho_0 \in C^\infty$ and decays exponentially fast as $|x| \to \infty$. Let ρ be the solution to the Fokker-Planck equation.

$$\frac{d}{dt}I(\rho|e^{-V}) = \frac{d}{dt}\int \rho|\nabla(\log \frac{\rho}{e^{-V}})|^2\, dx$$

Now, we perform a tedious calculation. For convenience, let us denote $\mu = e^{-V}$ and $q = \rho/e^{-V}$. This q also satisfies the backward equation.

$$\begin{aligned}\frac{d}{dt}\int \rho|\nabla \log q|^2\, dx &= \int \partial_t\rho|\nabla \log q|^2 + 2\int \rho\nabla \log q \nabla\frac{\partial_t q}{q} \\ &= \int \partial_t\rho|\nabla \log q|^2 - 2\int(\nabla\rho\nabla \log q + \rho\Delta \log q)\frac{\partial_t q}{q} \\ &= \int \partial_t\rho|\nabla \log q|^2 - 2\int(\mu\nabla q\cdot\nabla \log q - q\mu\nabla V\cdot\nabla \log q + \rho\Delta \log q)\frac{\partial_t q}{q} \\ &= -\int \partial_t\rho|\nabla \log q|^2 + 2\int \nabla V\cdot\nabla \log q\partial_t\rho - 2\int(\Delta \log q)\partial_t\rho =: -J\end{aligned}$$

Now, we know

$$\partial_t\rho = \nabla\cdot(e^{-V}\nabla(e^V\rho)),$$

then we can reduce the above to

$$\begin{aligned}J &= \int e^V\rho\nabla\cdot(e^{-V}\nabla(|\nabla \log q|^2 - 2\nabla V\cdot\nabla \log q + 2\Delta \log q))\, dx \\ &= -\int \rho\nabla V\cdot\nabla(|\nabla \log q|^2 - 2\nabla V\cdot\nabla \log q + 2\Delta \log q) + \int \rho\Delta(|\nabla \log q|^2 - 2\nabla V\cdot\nabla \log q + 2\Delta \log q)\end{aligned}$$

Now, consider

$$\int \rho(-2\Delta(\nabla V\cdot\nabla \log q))\, dx = 2\int \nabla\rho\cdot\nabla(\nabla V\cdot\nabla \log q) = 2\int \rho(\nabla \log q - \nabla V)\cdot\nabla(\nabla V\cdot\nabla \log q)$$

Adding this back, we have

$$J = -\int \rho\nabla V\cdot\nabla(|\nabla \log q|^2 + 2\Delta \log q) + \int \rho\Delta(|\nabla \log q|^2) + 2\int \Delta\rho\Delta \log q + 2\int \rho\nabla \log q\cdot\nabla(\nabla V\cdot\nabla \log q)$$

Note that

$$\nabla \log q\cdot\nabla(\nabla V\cdot\nabla \log q) = \nabla \log q\cdot\nabla^2 V\cdot\nabla \log q + \frac{1}{2}\nabla V\cdot\nabla|\nabla \log q|^2$$

we find

$$J = -2\int \rho\nabla V\cdot\nabla\Delta\log q + \int \rho\Delta(|\nabla\log q|^2) + 2\int \Delta\rho\Delta\log q + 2\int \rho\nabla\log q\cdot\nabla^2 V\cdot\nabla\log q$$

Now,

$$-2\int \rho\nabla V\cdot\nabla\Delta\log q + 2\int \Delta\rho\Delta\log q = -2\int(\rho\nabla V + \nabla\rho)\cdot\nabla\Delta\log q = -2\int \rho\nabla\log q\cdot\nabla\Delta\log q$$

Moreover, we have the identity

$$-\nabla u\cdot\nabla\Delta u + \Delta\frac{1}{2}|\nabla u|^2 = \nabla^2 u : \nabla^2 u.$$

Hence, we eventually find

$$J = 2\int \rho\nabla^2\log q : \nabla^2\log q + 2\int \rho\nabla\log q\cdot\nabla^2 V\cdot\nabla\log q \geq 2\lambda\int \rho|\nabla\log q|^2 = 2\lambda I(\rho|e^{-V})$$

This means

$$I(\rho|e^{-V}) \leq I(\rho_0|e^{-V})e^{-2\lambda t}$$

which implies that

$$H(\rho_0|e^{-V}) = \int_0^\infty I(\rho|e^{-V})dt \leq \frac{1}{2\lambda}I(\rho_0|e^{-V}).$$

If we recall ρ_0 is arbitrarily smooth functions that decay fast enough, by some density argument, this can be generalized to probability densities with bounded Fisher information.

Hence, we have established the Log Sobolev inequality. □

Remark 4. *The computation above is really mysterious. In fact, if we look at them in viewpoint of gradient flows in Wasserstein spaces, this will be very natrual by the uniform displacement convexity.*

As soon as we have the Log Sobolev inequality, we then have

$$H(\rho|e^{-V}) \leq H(\rho_0|e^{-V})e^{-2\lambda t}.$$

Finally, we use the Pinsker's inequality or Csiszar-Kullback inequality, we have

$$\|\rho - Ce^{-V}\|_1 \leq \sqrt{2H(\rho|Ce^{-V})}$$

where C is normalized so that Ce^{-V} is a probability density. This then shows the exponentially fast convergence of the probability distribution in total variation norm.

References

[1] R. Khasminskii. *Stochastic Stability of Differential Equations*, volume 66. Springer Science & Business Media, 2011.

[2] S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge University Press, Cambridge, second edition, 2009.

[3] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.*, 101(2):185–232, 2002.

[4] A. Eberle. Reflection coupling and Wasserstein contractivity without convexity. *Comptes Rendus Mathematique*, 349(19-20):1101–1104, 2011.

[5] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Relat. Fields*, 166(3-4):851–886, 2016.

[6] L. Gross. Existence and uniqueness of physical ground states. *Journal of Functional Analysis*, 10(1):52–109, 1972.

[7] F. Gong and L. Wu. Spectral gap of positive operators and applications. *Journal de Mathématiques Pures et Appliquées*, 85(2):151–191, 2006.

[8] L. Miclo. On hyperboundedness and spectrum of Markov operators. *Inventiones Mathematicae*, 200(1):311–343, 2015.

[9] P. A. Markowich and C. Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp*, 19:1–29, 2000.