Probabilistic Numerical Methods 2024–2025

Lecture 8: Long time behaviour

julien.reygner@enpc.fr

We fix b and σ which do not depend on time. We use the notation $\mathbb{P}_x$, $\mathbb{P}_\pi$ to say that $X_0 = x$ or $X_0 \sim \pi$.

1 Stationary distribution

1.1 Definition

A probability measure such that if $X_0 \sim \pi$ then $X_t \sim \pi$ for any $t \geq 0$.

π is stationary iff it is a distributional solution to the stationary Fokker–Planck equation $L^*\pi = 0$. As a consequence, if a smooth probability density is a classical solution to the PDE

$$0 = -\sum_{i=1}^{n} \partial_{x_i}(b_i(x)p(x)) + \frac{1}{2}\sum_{i,j=1}^{n} \partial_{x_i,x_j}(a_{ij}(x)p(x)).$$

Application: check that $\mathcal{N}(0, 1/2\lambda)$ is stationary for the Ornstein–Uhlenbeck process.

1.2 Reversibility

We say that the SDE is reversible with respect to π if $X_0 \sim \pi$ implies that for any $T > 0$, the processes $(X_t)_{t\in[0,T]}$ and $(X_{T-t})_{t\in[0,T]}$ have the same law. Then π is stationary.

Proposition 1.1 (Characterisation of reversibility)**.** *The SDE is reversible with respect to π if and only if for any f and g in C^2_c,*

$$\int_{x\in\mathbb{R}^n} Lf(x)g(x)\pi(\mathrm{d}x) = \int_{x\in\mathbb{R}^n} f(x)Lg(x)\pi(\mathrm{d}x).$$

An important application is the following. Let $V : \mathbb{R}^n \to \mathbb{R}$ and $\beta > 0$ such that

$$Z_\beta := \int_{x\in\mathbb{R}^n} \mathrm{e}^{-\beta V(x)}\mathrm{d}x < +\infty,$$

and define

$$\pi_\beta(\mathrm{d}x) = \frac{1}{Z_\beta}\mathrm{e}^{-\beta V(x)}\mathrm{d}x.$$

The (overdamped) *Langevin process* is the solution to the SDE

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t.$$

Lemme 1.2 (Reversibility of the Langevin process)**.** *The differential operator associated with the Langevin process writes*

$$Lf = \beta^{-1}\mathrm{e}^{\beta V}\nabla\cdot\left(\mathrm{e}^{-\beta V}\nabla f\right).$$

As a consequence, it is reversible with respect to π_β.

1.3 Existence by the Krylov–Bogoliubov Theorem

We give an existence result which is very similar to the one for Markov chains with finite state space. We first need to introduce some topological notions.

Definition 1.3 (Feller property)**.** *The semigroup $(P_t)_{t\geq 0}$ has the Feller property if for any $t \geq 0$ and for any bounded and continuous function $f : \mathbb{R}^n \to \mathbb{R}$, the function $P_t f : \mathbb{R}^n \to \mathbb{R}$ is continuous (and necessarily bounded).*

Definition 1.4 (Tightness)**.** *A family of probability measures $\{\mu_t, t > 0\}$ on $\mathbb{R}^n$ is tight if for any $\epsilon > 0$, there exist a compact set $K_\epsilon \subset \mathbb{R}^n$ such that for any $t > 0$, $\mu_t(K_\epsilon^c) \leq \epsilon$.*

Theorem 1.5 (Prohorov)**.** *If $\{\mu_t, t > 0\}$ is tight then it is relatively compact, namely any sequence $(\mu_{t_n})_{n\geq 1}$ admits a subsequence weakly converging toward some probability measure μ on $\mathbb{R}^n$.*

(The main reference is Billingsley – Convergence of Probability Measures).

We may now state the Krylov–Bogoliubov Theorem.

Theorem 1.6 (Krylov–Bogoliubov)**.** *Assume that the semigroup $(P_t)_{t\geq 0}$ has the Feller property and that there exists $x \in \mathbb{R}^n$ such that the family of probability measures $\{\widehat{\mu}_t, t > 0\}$ defined by*

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \qquad \widehat{\mu}_t(A) := \frac{1}{t} \int_{u=0}^{t} \mathbb{P}_x(X_u \in A)\mathrm{d}u$$

is tight. Then the semigroup $(P_t)_{t\geq 0}$ admits a stationary distribution.

Proof. By the tightness assumption and the Prohorov Theorem, there exist a sequence $t_n \to \infty$ and a probability measure π such that for any bounded and continuous function $f : \mathbb{R}^n \to \mathbb{R}$, $\widehat{\mu}_{t_n} f \to \pi f$. Now fix $s \geq 0$. By the Feller property, $\widehat{\mu}_{t_n} P_s f \to \pi P_s f$. But on the other hand, by the semigroup property,

$$\begin{aligned}
\widehat{\mu}_{t_n} P_s f &= \frac{1}{t_n} \int_{u=0}^{t_n} P_u(P_s f)(x)\mathrm{d}u \\
&= \frac{1}{t_n} \int_{u=0}^{t_n} P_{u+s} f(x)\mathrm{d}u \\
&= \frac{1}{t_n} \int_{u=s}^{s+t_n} P_v f(x)\mathrm{d}v \\
&= \widehat{\mu}_{t_n} f + \frac{1}{t_n}\left(-\int_{u=0}^{s} P_v f(x)\mathrm{d}v + \int_{u=t_n}^{s+t_n} P_v f(x)\mathrm{d}v\right),
\end{aligned}$$

and the term in parentheses is bounded by $2s\|f\|_\infty$. So $\widehat{\mu}_{t_n} P_s f \to \pi f$ and therefore π is stationary. □

To check the tightness condition one often uses a Lyapunov function.

Proposition 1.7 (Lyapunov condition)**.** *Assume that there is a C^2 function $\Phi : \mathbb{R}^n \to [0, +\infty)$ such that $\Phi(x) \to +\infty$ when $|x| \to +\infty$, and that for $c > 0$, $b \geq 0$,*

$$\forall x \in \mathbb{R}^n, \qquad L\Phi(x) \leq -c\Phi(x) + b.$$

1. *For any $x \in \mathbb{R}^n$ and $t \geq 0$, $\mathbb{E}_x[\Phi(X_t)] \leq \mathrm{e}^{-ct}\Phi(x) + b(1 - \mathrm{e}^{-ct})/c$.*
2. *For any choice of $x \in \mathbb{R}^n$, the family $\{\widehat{\mu}_t, t > 0\}$ defined in the Krylov–Bogoliubov Theorem is tight.*

3. *Any stationary distribution π given by the Krylov–Bogoliubov Theorem satisfies $\pi\Phi \leq b/c$.*

Proof. Applying Ito's formula and using the same localisation procedure as in the course on SDEs with locally Lipschitz coefficients, we get the first point. For the second point, let $M \geq 0$ and set $K = \{x \in \mathbb{R}^n : \Phi(x) \leq M\}$. Then for any $t \geq 0$,

$$\begin{aligned}
\widehat{\mu}_t(K^c) &= \frac{1}{t}\int_{u=0}^{t} \mathbb{P}_x(\Phi(X_u) > M)\mathrm{d}u \\
&\leq \frac{1}{t}\int_{u=0}^{t} \frac{\mathbb{E}_x[\Phi(X_u)]}{M}\mathrm{d}u \\
&\leq \frac{1}{M}(\Phi(x) + b/c),
\end{aligned}$$

where we have used the Markov inequality at the second line and the first part of the proposition at the third line. For fixed ϵ, M can be chosen large enough for the right-hand side to be smaller than ϵ, which shows the second point. The third point is left as an exercise. □

1.4 Uniqueness, irreducibility and ergodicity

We give a rather strong irreducibility condition.

Proposition 1.8 (Irreducibility condition)**.** *Assume that there is $t_0 > 0$ such that the transition kernel $p_{t_0}(x, \mathrm{d}y)$ has a continuous (in y) density $p_{t_0}(x, y)$, and that there is a nonempty open set $B \subset \mathbb{R}^n$ such that*

$$\forall x \in \mathbb{R}^n, \quad \forall y \in B, \qquad p_{t_0}(x, y) > 0.$$

Then $(P_t)_{t\geq 0}$ has at most one stationary distribution.

The irreducibility condition is typically true if b and σ are smooth and $a = \sigma\sigma^\top$ is uniformly elliptic, that is to say that there exist $\lambda > 0$ such that for any $x, \xi \in \mathbb{R}^n$, $\xi \cdot a(x)\xi \geq \lambda|\xi|^2$. From a probabilistic point of view this says that the Brownian motion can 'diffuse in all directions'.

To understand why Proposition 1.8 holds true, assume that $(P_t)_{t\geq 0}$ has the Feller property and denote by $\mathcal{I}$ the set of stationary distributions for the SDE. This is a convex, closed subset of the set of probability measures on $\mathbb{R}^n$ (endowed with the topology of weak convergence). The extremal points of $\mathcal{I}$, that is to say the points which can not be written as convex combinations of other points of $\mathcal{I}$, are called *ergodic*. Then we have the two statements:

1. Two distinct ergodic measures are mutually singular.
2. Birkhoff's Ergodic Theorem: for any ergodic measure π, for π-almost every $x \in \mathbb{R}^n$, for any $f \in \mathbf{L}^1(\pi)$,

$$\lim_{t\to+\infty} \frac{1}{t}\int_{s=0}^{t} f(X_s)\mathrm{d}s = \pi f, \qquad \mathbb{P}_x\text{-almost surely.} \tag{1}$$

Under the irreducibility condition of Proposition 1.8, we therefore deduce that any stationary distribution π satisfies

$$\pi(B) = \mathbb{P}_\pi(X_{t_0} \in B) = \int_{x\in\mathbb{R}^n} \pi(\mathrm{d}x) \int_{y\in B} p_{t_0}(x, y)\mathrm{d}y > 0.$$

This implies that there is at most one ergodic measure, and then at most one stationary distribution. See Section 3.1 in https://doi.org/10.1051/ps:2001106 for details and further references.

2 LLN and CLT

If the conditions of Propositions 1.7 and 1.8 hold, then there is a unique stationary distribution π. Besides, it may be shown (still Section 3.1 in https://doi.org/10.1051/ps:2001106) that the convergence (1) actually holds for *all* initial conditions $x \in \mathbb{R}^n$. This is the continuous state space equivalent of the notion of positive recurrence.

The next question is whether this LLN result may be complemented by a CLT: in other words, does

$$\sqrt{t}\left(\frac{1}{t}\int_{s=0}^{t} f(X_s)\mathrm{d}s - \pi f\right)$$

converge in distribution to some Gaussian measure?

2.1 CLT by Poisson equation

To study this question, let us assume that we can find a 'nice' solution to the Poisson equation

$$Lg = f - \pi f \tag{2}$$

on $\mathbb{R}^n$. The precise meaning of 'nice' will be clarified below. Formally, we should think of g as

$$L^{-1}(f - \pi f) = -\int_{t=0}^{+\infty} \mathrm{e}^{tL}(f - \pi f)\mathrm{d}t = -\int_{t=0}^{+\infty}(P_t f - \pi f)\mathrm{d}t,$$

provided that the right-hand side makes sense in an appropriate functional space. By Ito's formula we thus have

$$g(X_t) = g(X_0) + \int_{s=0}^{t} Lg(X_s)\mathrm{d}s + \int_{s=0}^{t} \sigma^\top(X_s)\nabla g(X_s)\cdot \mathrm{d}B_s,$$

which rewrites

$$\sqrt{t}\left(\frac{1}{t}\int_{s=0}^{t} f(X_s)\mathrm{d}s - \pi f\right) = \frac{g(X_t) - g(X_0)}{\sqrt{t}} - \frac{1}{\sqrt{t}}\int_{s=0}^{t}\sigma^\top(X_s)\nabla g(X_s)\cdot \mathrm{d}B_s$$

by (2). Assume that $g(X_t)/\sqrt{t} \to 0$ in probability, so we only have to care about the stochastic integral. By the Dambis–Dubins–Schwartz Theorem, there is a (scalar) Brownian motion $(\beta_r)_{r\geq 0}$ such that

$$\frac{1}{\sqrt{t}}\int_{s=0}^{t}\sigma^\top(X_s)\nabla g(X_s)\cdot \mathrm{d}B_s = \frac{1}{\sqrt{t}}\beta_{\int_{s=0}^{t}|\sigma^\top(X_s)\nabla g(X_s)|^2\mathrm{d}s}.$$

A (a bit bold[1]) application of the scale invariance of the Brownian motion then indicates that

$$\frac{1}{\sqrt{t}}\beta_{\int_{s=0}^{t}|\sigma^\top(X_s)\nabla g(X_s)|^2\mathrm{d}s} \overset{\mathcal{L}}{=} \beta_{\frac{1}{t}\int_{s=0}^{t}|\sigma^\top(X_s)\nabla g(X_s)|^2\mathrm{d}s},$$

and since by the ergodic theorem,

$$\lim_{t\to+\infty}\frac{1}{t}\int_{s=0}^{t}|\sigma^\top(X_s)\nabla g(X_s)|^2\mathrm{d}s = \int_{x\in\mathbb{R}^n}|\sigma^\top(x)\nabla g(x)|^2\pi(\mathrm{d}x) =: \sigma^2(f), \qquad \text{almost surely,}$$

we conclude that

$$\lim_{t\to+\infty}\sqrt{t}\left(\frac{1}{t}\int_{s=0}^{t} f(X_s)\mathrm{d}s - \pi f\right) = \beta_{\sigma^2(f)} \sim \mathcal{N}(0,\sigma^2(f)), \qquad \text{in distribution.}$$

So in the end, a 'nice' solution g to (2) should:

[1] This step is more conventionally handled by martingale convergence theorems.

- be C^2 (for the Ito formula to be applied),
- be such that $g(X_t)/\sqrt{t} \to 0$ in probability,
- be such that $\frac{1}{t}\int_{s=0}^{t} |\sigma^\top(X_s)\nabla g(X_s)|^2 \mathrm{d}s \to \int_{x\in\mathbb{R}^n} |\sigma^\top(x)\nabla g(x)|^2 \pi(\mathrm{d}x) < +\infty$.

Under these conditions, we get the expected CLT.

2.2 Alternative formula for $\sigma^2(f)$

To get a more explicit formula for $\sigma^2(f)$ we first introduce the *carré du champ* operator Γ.

Definition 2.1 (Carré du champ). *For any C^1 functions $f, g : \mathbb{R}^n \to \mathbb{R}$, let us define*

$$\Gamma(f,g)(x) = \frac{1}{2}(\sigma^\top(x)\nabla f(x)) \cdot (\sigma^\top(x)\nabla g(x)).$$

When f and g are C^2 this operator satisfies

$$\Gamma(f,g) = \frac{1}{2}\left(L(fg) - fLg - gLf\right).$$

It is called the carré du champ.

The asymptotic variance in the CLT then rewrites

$$\sigma^2(f) = 2\int \Gamma(g,g)\pi.$$

Using the expression of $\Gamma(g,g)$ in terms of L, we get

$$\sigma^2(f) = \int (Lg^2 - 2gLg)\pi.$$

Since $L^*\pi = 0$, the first term vanishes, while using the Poisson equation and the expression of g in terms of f,

$$\begin{aligned}
\sigma^2(f) &= -2\int gLg)\pi \\
&= 2\int \left(\int_{t=0}^{\infty} (P_t f - \pi f)\mathrm{d}t\right)(f - \pi f)\pi \\
&= 2\int_{t=0}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_t))\mathrm{d}t.
\end{aligned}$$

3 Convergence to equilibrium

For the expressions of the asymptotic variance in the CLT derived above to make sense, it is necessary that $P_t f$ converges to πf fast enough. Recalling that $P_t f(x) = \mathbb{E}_x[f(X_t)]$, this basically means that X_t must converge to π in distribution. So our goal here is either to quantify the convergence of $\mathrm{Law}(X_t)$ to π for a given distance on the set of probability measures, or of the function $P_t f$ toward the constant πf in a given functional space.

3.1 Probabilistic construction and couplings

3.1.1 Total variation estimate

We start with a construction which relies on the Lyapunov condition of Proposition 1.7 combined with a quantitative version of the irreducibility condition of Proposition 1.8: we assume that for any compact subset $K \subset \mathbb{R}^n$, there exist $t_0 > 0$, $\epsilon > 0$ and a probability measure ν on $\mathbb{R}^n$ such that

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \quad \forall x \in K, \qquad \mathbb{P}_x(X_{t_0} \in A) \geq \epsilon\nu(A).$$

This condition is sometimes called the *Doeblin*, or *minorisation* condition. Then, given two probability measures μ and ν on $\mathbb{R}^n$, it is possible to construct a pair of processes $(X_t, Y_t)_{t\geq 0}$ such that $X_0 \sim \mu$, $Y_0 \sim \nu$, both $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are solution to the SDE (with different driving Brownian motions), and:

- the time for (X_t, Y_t) to enter $K \times K$ is controlled by the Lyapunov condition;
- if $(X_t, Y_t) \in K \times K$, then $X_{t+t_0} = Y_{t+t_0}$ with probability at least ϵ.

Once the coupling has succeeded, we let X_t and Y_t evolve together. Denoting by τ the coupling time, we therefore have that τ is dominated by a random variable of the form κT, with κ a geometric variable with parameter ϵ. Let us then introduce the *total variation* between two probability measures μ and ν by any of the three equivalent formulas:

$$\begin{aligned}
\|\mu - \nu\|_{\mathrm{TV}} &= \max_{A\in\mathcal{B}(\mathbb{R}^n)} |\mu(A) - \nu(A)| \\
&= \sup_{\|f\|_\infty \leq 1} |\mu f - \nu f| \\
&= \inf_{X\sim\mu, Y\sim\nu} \mathbb{P}(X \neq Y),
\end{aligned}$$

where in the last line, the infimum is taken over all pairs of random variables (X, Y) with marginal distributions μ and ν. Denoting by μ_t the law of X_t and by ν_t the law of Y_t, we deduce from the third expression that

$$\|\mu_t - \nu_t\|_{\mathrm{TV}} \leq \mathbb{P}(\tau > t) \leq C\mathrm{e}^{-\beta t}$$

for some $C \geq 0$ and $\beta > 0$ which depend on T and ϵ. In particular, if $\nu = \pi$ then $\nu_t = \pi$ for any t, and therefore we get exponential convergence to π.

Details on this approach can be found in the lecture notes http://www.hairer.org/notes/Convergence.pdf.

3.1.2 Wasserstein estimate

The previous approach aims at constructing a coupling of two solutions to the SDE which coincide after an almost surely finite time. One may be slightly less ambitious and only try to construct two solutions which 'get closer' to each other. In this case, the appropriate notion of distance to measure how far μ_t and ν_t are is the *Wasserstein distance* defined, for $p \geq 1$, by

$$W_p(\mu, \nu) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[|X - Y|^p]^{1/p}.$$

Consider for instance two solutions $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ driven by *the same Brownian motion* $(B_t)_{t\geq 0}$ (this coupling is called *synchronous*):

$$\begin{aligned}
\mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad X_0 \sim \mu, \\
\mathrm{d}Y_t &= b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}B_t, \qquad Y_0 \sim \nu.
\end{aligned}$$

By Ito's formula,

$$\mathrm{d}|X_t - Y_t|^2 = 2(X_t - Y_t)\cdot((b(X_t) - b(Y_t))\mathrm{d}t + (\sigma(X_t) - \sigma(Y_t))\mathrm{d}B_t) + |\sigma(X_t) - \sigma(Y_t)|^2\mathrm{d}t,$$

so that assuming that the stochastic integral vanishes when taking the expectation, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[|X_t - Y_t|^2\right] = \mathbb{E}\left[2(X_t - Y_t)\cdot(b(X_t) - b(Y_t)) + |\sigma(X_t) - \sigma(Y_t)|^2\right].$$

As a conclusion, if b and σ satisfy the condition that there exists $c > 0$ such that

$$\forall x, y \in \mathbb{R}^n, \qquad 2(x-y)\cdot(b(x) - b(y)) + |\sigma(x) - \sigma(y)|^2 \leq -2c|x-y|^2, \tag{3}$$

we deduce that

$$\mathbb{E}\left[|X_t - Y_t|^2\right] \leq \mathrm{e}^{-2ct}\mathbb{E}\left[|X_0 - Y_0|^2\right].$$

The left-hand side is larger than $W_2(\mu_t, \nu_t)^2$, and then taking the infimum over all couplings (X_0, Y_0) of μ, ν, we deduce that

$$W_2(\mu_t, \nu_t) \leq \mathrm{e}^{-ct} W_2(\mu, \nu).$$

Once again, if we take $\nu = \pi$ we get exponential convergence to π.

The condition (3) may look rather 'ad hoc'. It holds in particular for the Langevin process $\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t$ when the potential function V is uniformly convex. Beyond this convex case, other couplings may be employed, such as reflection couplings (see http://www.numdam.org/articles/10.1016/j.crma.2011.09.003/).

3.2 Spectral approach and functional inequalities

Since $\mathrm{Law}(X_t)$ and $P_t f$ are solutions to linear evolutionary PDE, their long time behaviour can also be studied by PDE/analytic techniques. At least in the reversible case, where L and P_t are symmetric in $\mathbf{L}^2(\pi)$, functional analysis allows to study the convergence of $P_t f = \mathrm{e}^{tL} f$ to πf in a rather abstract setting (see Section 3.2 of https://doi.org/10.1051/proc/201444006). We first present the idea of this approach in a simplified setting, and then explain how functional inequalities may be used as a substitute to this abstract setting.

3.2.1 Spectral approach in a caricatural setting

In the reversible case, the basic remark is that the operator L (which is symmetric in $\mathbf{L}^2(\pi)$) has eigenvalue 0, with associated eigenspace the space of constant functions $\mathbb{R}^n \to \mathbb{R}$, and it is nonpositive – indeed, for any smooth f,

$$\int_{x\in\mathbb{R}^n} f(x)Lf(x)\pi(\mathrm{d}x) = -\int_{x\in\mathbb{R}^n} \Gamma(f,f)(x)\pi(\mathrm{d}x) \leq 0.$$

So let us transpose these properties in a finite-dimensional setting: let E be a finite space with cardinality m, π be a probability measure on E (with $\pi(x) > 0$ for all $x \in E$). We denote by $\ell^2(\pi)$ the set of functions $E \to \mathbb{R}$ equipped with the scalar product

$$\langle f, g\rangle_\pi = \sum_{x\in E} f(x)g(x)\pi(x).$$

Let L be an $\mathsf{m}\times\mathsf{m}$ matrix which is symmetric in $\ell^2(\pi)$, with eigenvalues $0 = \lambda_1 > -\lambda_2 \geq \cdots \geq -\lambda_{\mathsf{m}}$, and eigenvectors $e_1, \ldots, e_{\mathsf{m}}$ which form an orthonormal basis of $\ell^2(\pi)$, with $e_1 = \mathbf{1}$ (the vector with all entries equal to 1). The assumption that $\lambda_2 > 0$, and therefore that 0 is a simple eigenvalue, corresponds to π being the only stationary distribution. We then define

$$\forall f \in \ell^2(\pi), \quad \forall t \geq 0, \qquad P_t f = \mathrm{e}^{tL} f = \sum_{i=1}^{\mathsf{m}} \mathrm{e}^{-\lambda_i t} \langle f, e_i \rangle_\pi e_i.$$

Remark 3.1. *This setting is, in fact, not artificial at all. If we let $\widetilde{P}$ be a stochastic matrix on E (irreducible, with unique stationary distribution π, and reversible with respect to π), let $(\widetilde{X}_n)_{n\geq 0}$ be a (discrete time) Markov chain with transition matrix $\widetilde{P}$, consider an independent sequence of random times $(T_n)_{n\geq 0}$ with $T_0 = 0$ and $(T_n - T_{n-1})_{n\geq 1}$ independent with Exponential distribution with parameter 1, and finally set $X_t = \widetilde{X}_n$ on $[T_n, T_{n+1})$, then it turns out that $(X_t)_{t\geq 0}$ is a continuous time Markov process, with infinitesimal generator L and semigroup $(P_t)_{t\geq 0}$. So what we are describing here is exactly the spectral theory of its long time convergence.*

The constant function πf is exactly $\langle f, e_1 \rangle_\pi e_1$, so we have

$$P_t f - \pi f = \sum_{i=2}^{\mathsf{m}} \mathrm{e}^{-\lambda_i t} \langle f, e_i \rangle_\pi e_i,$$

and therefore

$$\begin{aligned} \|P_t f - \pi f\|_\pi^2 &= \sum_{i=2}^{\mathsf{m}} \mathrm{e}^{-2\lambda_i t} |\langle f, e_i \rangle_\pi|^2 \\ &\leq \mathrm{e}^{-2\lambda_2 t} \sum_{i=2}^{\mathsf{m}} |\langle f, e_i \rangle_\pi|^2 \\ &= \mathrm{e}^{-2\lambda_2 t} \|f - \pi f\|_\pi^2. \end{aligned}$$

This shows that the decay of $P_t f - \pi f$ is governed by λ_2, the smallest nonzero eigenvalue of $-L$, which is called the *spectral gap* of L. Our goal is now to give a characterisation of this quantity which does not involve the spectrum of L, so as to be generalised to the SDE case without heavy functional analysis. To proceed, let us introduce the *Dirichlet form*

$$\forall f \in \ell^2(\pi), \qquad \mathcal{E}(f) = -\langle f, Lf \rangle_\pi = \sum_{i=1}^{\mathsf{m}} \lambda_i |\langle f, e_i \rangle_\pi|^2.$$

Since $\lambda_1 = 0$, it is clear that we have

$$\lambda_2 = \inf_{\mathrm{Var}_\pi(f) > 0} \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)},$$

where we have introduced the notation $\mathrm{Var}_\pi(f) = \|f - \pi f\|_\pi^2$. So let us summarise: with the spectral gap λ_2 defined in terms of the functionals $\mathcal{E}$ and Var_π, we have the exponential decay

$$\forall t \geq 0, \qquad \mathrm{Var}_\pi(P_t f) \leq \mathrm{e}^{-2\lambda_2 t} \mathrm{Var}_\pi(f). \tag{4}$$

3.2.2 Poincaré inequality

We now aim at formulating the same statement coming back to the SDE case, and without speaking of the spectrum of L – in fact, without even assuming the reversibility of the process. Let us first define the functionals

$$\begin{aligned}
\mathrm{Var}_\pi(f) &= \int_{x\in\mathbb{R}^n} (f(x)-\pi f)^2 \pi(\mathrm{d}x),\\
\mathcal{E}(f) &= -\int_{x\in\mathbb{R}^n} f(x) Lf(x)\pi(\mathrm{d}x) = \int_{x\in\mathbb{R}^n} \Gamma(f,f)(x)\pi(\mathrm{d}x),
\end{aligned}$$

and then the spectral gap $\lambda_2 \geq 0$ by

$$\lambda_2 = \inf_{\mathrm{Var}_\pi(f)>0} \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)}.$$

Notice that any $c \geq 1/\lambda_2$ satisfies

$$\forall f, \qquad \mathrm{Var}_\pi(f) \leq c\mathcal{E}(f),$$

which rewrites explicitly

$$\forall f, \qquad \int_{x\in\mathbb{R}^n} (f(x)-\pi f)^2 \pi(\mathrm{d}x) \leq c \int_{x\in\mathbb{R}^n} |\sigma^\top(x)\nabla f(x)|^2 \pi(\mathrm{d}x).$$

Seeing the left-hand side as an $\mathbf{L}^2$ norm and the right-hand side as an $\mathbf{H}^1$ norm, we recognise a *Poincaré inequality*. Then the convergence statement (4) may be recovered as follows.

Proposition 3.2 (Exponential convergence under Poincaré inequality)**.** *Assume that the Poincaré inequality is satisfied with a constant c. Then, for any f,*

$$\forall t \geq 0, \qquad \mathrm{Var}_\pi(P_t f) \leq \mathrm{e}^{-2t/c}\,\mathrm{Var}_\pi(f).$$

Proof. We just compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Var}_\pi(P_t f) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{x\in\mathbb{R}^n} (P_t f(x)-\pi f)^2 \pi(\mathrm{d}x)\\
&= 2\int_{x\in\mathbb{R}^n} \frac{\mathrm{d}}{\mathrm{d}t} P_t f(x)(P_t f(x)-\pi f)\pi(\mathrm{d}x)\\
&= 2\int_{x\in\mathbb{R}^n} LP_t f(x)(P_t f(x)-\pi f)\pi(\mathrm{d}x)\\
&= -2\int_{x\in\mathbb{R}^n} \Gamma(P_t f, P_t f)(x)\pi(\mathrm{d}x) = -2\mathcal{E}(P_t f),
\end{aligned}$$

and we conclude using the Poincaré inequality. □

An important remark here is that this computation does not require reversibility, nor spectral analysis of L.

The Poincaré inequality is just an example of the use of functional inequalities to study the long time behaviour of stochastic processes. To learn more about this topic and its applications to numerical methods in statistical physics, you are strongly encouraged to take the course *Computational Statistical Physics* next semester!