On the exponential convergence of Langevin diffusions: from deterministic to stochastic stability

Stephen Tu
Google Brain Robotics

June 17, 2022

Abstract

Consider a dynamical system described by the ordinary differential equation $dX/dt = -\nabla U(X)$, and its corresponding stochastic differential equation:

$$dX = -\nabla U(X)dt + \sqrt{2}dB.$$

Under what conditions does stability of the deterministic dynamics $dX/dt = -\nabla U(X)$ imply convergence of the stochastic dynamics to its stationary distribution $\mu = \exp(-U)$? In this expository paper, we give a self-contained derivation of the basic results addressing this question. We first show how both the Poincaré and log-Sobolev functional inequalities naturally arise from the desire of exponential convergence. We then construct a stochastic Lyapunov criteria for certifying the Poincaré inequality, and then show how classical Lyapunov analysis on $dX/dt = -\nabla U(X)$ directly translates to stochastic Lyapunov functions. While this paper contains no new results, all proofs given are direct and elementary.

1 Introduction and problem formulation

Let μ be a probability measure on $\mathbb{R}^n$ with density (w.r.t. the Lebesgue measure) given by $\exp(-U)$, where $U : \mathbb{R}^n \to \mathbb{R}$ is a smooth potential function. We consider the Langevin diffusion:

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dB_t. \tag{1.1}$$

Here, (B_t) is standard Brownian motion in $\mathbb{R}^n$. The main question addressed in this paper is the following:

When does the stability of the deterministic dynamics $dX/dt = -\nabla U(X)$ imply convergence of the stochastic differential equation (1.1) to the stationary distribution μ?

The study of diffusions of the form (1.1) has a rich history in the literature—see e.g. Bakry et al. [2014] for a comprehensive treatment of the subject. As such, it is not surprising that the answer to our question can already be pieced together from existing results. The purpose of this expository paper is to give a self-contained treatment of these results in the most direct and elementary manner possible.

The main result in this paper is Theorem 4.3, described in full detail in Section 4. At an informal level, it states that if the dynamical system $dX/dt = -\nabla U(X)$ is exponentially stable, then the Lyapunov function which certifies its exponential stability can be used to also certify that the distribution of X_t in the diffusion (1.1) converges exponentially fast to the stationary distribution μ.

The path we take towards proving the main result is fairly standard. It is well-known that if the stationary distribution μ satisfies a functional inequality such as the Poincaré or log-Sobolev inequality, then the distribution of X_t in (1.1) converges exponentially fast to μ in the χ^2-divergence (for Poincaré) or the KL-divergence (for log-Sobolev). Section 2 gives a self-contained proof of these facts using only basic tools. Next, in Section 3, we use the work of Bakry et al. [2008] to construct a stochastic Lyapunov criteria that establishes the Poincaré inequality. Remarkably, this sufficient condition also has an elementary proof. Finally, we bring all the pieces together by translating deterministic Lyapunov functions into stochastic Lyapunov functions in Section 4, yielding the main result Theorem 4.3.

1.1 Preliminaries

Notation For a vector $x \in \mathbb{R}^n$, the norms $\|x\|$ and $\|x\|_\infty$ indicate the ℓ_2 and ℓ_∞ norms in $\mathbb{R}^n$, respectively. The closed ℓ_2 ball of $\mathbb{R}^n$ with radius R is denoted $B_2^n(R)$. For a function $f : \mathbb{R}^n \to \mathbb{R}$, ∇f denotes its gradient and $\Delta f = \mathrm{tr}(\nabla^2 f)$ denotes its Laplacian. For a vector field $F : \mathbb{R}^n \to \mathbb{R}^n$, $\nabla \cdot F = \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}$ denotes its divergence.

Divergences and entropy For two distributions μ, ν on a common measurable space $(\Omega, \mathscr{A})$ with $\mu \ll \nu$, the χ^2-divergence and KL-divergences are defined as:

$$\chi^2(\mu, \nu) := \mathrm{Var}_\nu\left(\frac{d\mu}{d\nu}\right), \quad \mathsf{KL}(\mu, \nu) = \mathbb{E}_\mu \log \frac{d\mu}{d\nu}.$$

The following relationship between χ^2 and KL-divergence holds [Tsybakov, 2009, Lemma 2.7]:

$$\mathsf{KL}(\mu, \nu) \leqslant \log(1 + \chi^2(\mu, \nu)) \leqslant \chi^2(\mu, \nu). \tag{1.2}$$

Both these divergences upper bound the total-variation distance $\|\mu - \nu\|_{\mathrm{tv}} := \sup_{A \in \mathscr{A}} |\mu(A) - \nu(A)|$ via Pinsker's inequality [Tsybakov, 2009, Lemma 2.5]:

$$\|\mu - \nu\|_{\mathrm{tv}} \leqslant \sqrt{\mathsf{KL}(\mu, \nu)/2}.$$

For a measure μ and non-negative random variable X, the entropy of X is defined as:

$$\mathrm{Ent}_\mu(X) := \mathbb{E}_\mu[X \log X] - (\mathbb{E}_\mu X) \log \mathbb{E}_\mu X.$$

Functional inequalities The measure μ satisfies the *Poincaré inequality* with constant C_{PI} if for all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$:

$$\mathrm{Var}_\mu(f) \leqslant C_{\mathsf{PI}} \mathbb{E}_\mu \|\nabla f\|^2. \tag{1.3}$$

On the other hand, the measure μ satisfies the *log-Sobolev inequality* (LSI) with constant C_{LSI} if for all smooth functions f:

$$\mathrm{Ent}_\mu(f^2) \leqslant 2C_{\mathsf{LSI}} \mathbb{E}_\mu \|\nabla f\|^2. \tag{1.4}$$

The log-Sobolev inequality is stronger than the Poincaré inequality: if a measure μ satisfies the log-Sobolev inequality with constant C, then it also satisfies the Poincaré inequality with the same constant. We provide a self-contained proof of this well-known fact in Appendix A.

Integration by parts We will commonly use the following integration by parts identities. Let f, g map $\mathbb{R}^n \to \mathbb{R}$ and let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Assume that these functions are sufficiently smooth, and the behavior of these functions at infinity is zero. Integration by parts yields:

$$\int \langle F, \nabla f \rangle dx = -\int (\nabla \cdot F) f dx, \quad \int f \Delta g dx = -\int \langle \nabla f, \nabla g \rangle dx.$$

2 Functional inequalities imply exponential convergence

In this section, we will show that both the Poincaré inequality and log-Sobolev inequality give rise to exponential convergence of the diffusion (1.1) to the stationary distribution μ. The main difference is the divergence in which the convergence is measured.

Let μ_t denote the density of X_t w.r.t. the Lebesgue measure.[1] By the Fokker-Planck equation:

$$\partial_t \mu_t = \nabla \cdot (\mu_t \nabla U) + \Delta \mu_t, \tag{2.1}$$

where the boundary condition imposed on μ_0 encodes the distribution used to initialize X_0. While (2.1) is the standard form of the Fokker-Planck equation, for the proofs we will give, it is more natural to express it involving μ_t and μ instead of μ_t and ∇U as follows:

$$\begin{aligned}
\partial_t \mu_t &= \nabla \cdot (\mu_t \nabla U) + \Delta \mu_t \\
&= \nabla \cdot \left(\mu_t \left[\nabla U + \frac{\nabla \mu_t}{\mu_t} \right] \right) && \text{since } \Delta \mu_t = \nabla \cdot \nabla \mu_t \\
&= \nabla \cdot (\mu_t \nabla \log \mu_t - \mu_t \nabla \log \mu) && \text{since } \nabla \log \mu_t = \frac{\nabla \mu_t}{\mu_t} \text{ and } \mu = \exp(-U) \\
&= \nabla \cdot \left(\mu_t \nabla \log \frac{\mu_t}{\mu} \right). && (2.2)
\end{aligned}$$

Observe that applying (2.2) to μ yields $\partial_t \mu = 0$, confirming that μ is a stationary distribution of (1.1).

For what follows, all proofs will be informal in the sense that we will not check the various technical conditions necessary to interchange order of differentiation and integration, apply integration by parts, etc. All these proofs can be made rigorous by standard arguments. The following proposition shows that the Poincaré inequality implies exponential convergence to μ in the χ^2-distance.

Lemma 2.1. *Suppose μ satisfies the Poincaré inequality with constant C_{PI}. Then for any initial μ_0:*

$$\chi^2(\mu_t, \mu) \leqslant \exp\left(-\frac{2t}{C_{\mathsf{PI}}} \right) \chi^2(\mu_0, \mu).$$

Proof. By the Poincaré inequality (1.3) applied to $f = \mu_t / \mu$:

$$\chi^2(\mu_t, \mu) = \mathrm{Var}_\mu \left(\frac{\mu_t}{\mu} \right) \leqslant C_{\mathsf{PI}} \mathbb{E}_\mu \left\| \nabla \left(\frac{\mu_t}{\mu} \right) \right\|^2. \tag{2.3}$$

Therefore:

$$\frac{d}{dt} \chi^2(\mu_t, \mu) = \frac{d}{dt} \int \frac{(\mu_t - \mu)^2}{\mu} dx$$

[1] We will often use the same notation to refer to a measure and its density.

$$
\begin{aligned}
&= 2\int \frac{\mu_t - \mu}{\mu} \partial \mu_t dx \\
&= 2\int \frac{\mu_t - \mu}{\mu} \nabla \cdot \left(\mu_t \nabla \log \frac{\mu_t}{\mu} \right) dx && \text{Fokker-Planck equation (2.2)} \\
&= -2\int \left\langle \mu_t \nabla \log \frac{\mu_t}{\mu}, \nabla \left(\frac{\mu_t}{\mu} \right) \right\rangle dx && \text{integration by parts} \\
&= -2\int \left\| \nabla \left(\frac{\mu_t}{\mu} \right) \right\|^2 \mu dx && \text{since } \mu_t \nabla \log \frac{\mu_t}{\mu} = \mu \nabla \left(\frac{\mu_t}{\mu} \right) \\
&\leqslant -\frac{2}{C_{\mathsf{PI}}} \chi^2(\mu_t, \mu) && \text{using (2.3).}
\end{aligned}
$$

The claim now follows by the comparison lemma. □

The next lemma parallels Lemma 2.1, but uses the log-Sobolev inequality instead. The main difference is that convergence is given in the KL-divergence.

Lemma 2.2. *Suppose μ satisfies the log-Sobolev inequality with constant C_{LSI}. Then for any initial μ_0:*

$$
\mathsf{KL}(\mu_t, \mu) \leqslant \exp\left(-\frac{2t}{C_{\mathsf{LSI}}} \right) \mathsf{KL}(\mu_0, \mu).
$$

Proof. This proof is from Vempala and Wibisono [2019, Lemma 2]. First, note that:

$$
\int \mu_t \partial_t \log \mu_t dx = \int \partial_t \mu_t dx = \partial_t \int \mu_t dx = 0. \tag{2.4}
$$

Next, it is straightforward to verify:

$$
\mathbb{E}_{\mu_t} \left\| \nabla \log \frac{\mu_t}{\mu} \right\|^2 = 4\mathbb{E}_{\mu} \left\| \nabla \sqrt{\frac{\mu_t}{\mu}} \right\|^2. \tag{2.5}
$$

Applying the log-Sobolev inequality (1.4) to the function $f = \sqrt{\mu_t/\mu}$ and using (2.5):

$$
\mathsf{KL}(\mu_t, \mu) = \mathrm{Ent}_{\mu} \left(\frac{\mu_t}{\mu} \right) \leqslant 2C_{\mathsf{LSI}} \mathbb{E}_{\mu} \left\| \nabla \sqrt{\frac{\mu_t}{\mu}} \right\|^2 = \frac{C_{\mathsf{LSI}}}{2} \mathbb{E}_{\mu_t} \left\| \nabla \log \frac{\mu_t}{\mu} \right\|^2. \tag{2.6}
$$

Hence:

$$
\begin{aligned}
\frac{d}{dt} \mathsf{KL}(\mu_t, \mu) &= \frac{d}{dt} \int \mu_t \log \frac{\mu_t}{\mu} dx \\
&= \int \partial_t \mu_t \log \frac{\mu_t}{\mu} dx + \int \mu_t \partial_t \log \mu_t dx \\
&= \int \partial_t \mu_t \log \frac{\mu_t}{\mu} dx && \text{since } \int \mu_t \partial_t \log \mu_t dx = 0 \text{ by (2.4)} \\
&= \int \log \frac{\mu_t}{\mu} \nabla \cdot \left(\mu_t \nabla \log \frac{\mu_t}{\mu} \right) dx && \text{Fokker-Planck equation (2.2)} \\
&= -\int \left\| \nabla \log \frac{\mu_t}{\mu} \right\|^2 \mu_t dx && \text{integration by parts} \\
&\leqslant -\frac{2}{C_{\mathsf{LSI}}} \mathsf{KL}(\mu_t, \mu) && \text{using (2.6).}
\end{aligned}
$$

The claim now follows by the comparison lemma. □

At this point, the attentive reader will notice an apparent paradox. As claimed in Section 1 and shown in Proposition A.1, the log-Sobolev inequality implies the Poincaré inequality. Furthermore, the convergence result under the LSI (Lemma 2.2) yields convergence in the KL-divergence, compared with convergence in χ^2-divergence given in Lemma 2.1 under the Poincaré inequality. However, the KL-divergence is dominated by the χ^2-divergence via (1.2). So at first glance, it appears that the LSI assumption yields a weaker convergence bound.

The key to resolving this is to observe that the divergence between μ_0 and μ at initialization matters. Suppose for simplicity that both μ_0 and μ are Gaussian measures. Then, generically the KL-divergence $\mathsf{KL}(\mu_0, \mu)$ scales as order n. However, the bound $\mathsf{KL}(\mu_0, \mu) \leqslant \log(1 + \chi^2(\mu_0, \mu))$ implies that generically, $\chi^2(\mu_0, \mu)$ must be of order e^n. Thus, in order to reach $\chi^2(\mu_t, \mu) \leqslant \varepsilon$, Lemma 2.1 states that $t \asymp C_{\mathsf{PI}}(n + \log(1/\varepsilon))$ time suffices. On the other hand, $\mathsf{KL}(\mu_t, \mu) \leqslant \varepsilon$ is reached in time $t \asymp C_{\mathsf{LSI}} \log(n/\varepsilon)$ by Lemma 2.2.

The difference in convergence rate between the Poincaré and log-Sobolev inequality can be made more transparent by considering the q-Rényi divergence:

$$R_q(\mu, \nu) := \frac{1}{q-1} \log \mathbb{E}_\nu \left(\frac{d\mu}{d\nu} \right)^q, \ \ q > 0, \ \ q \neq 1.$$

The relationship $R_2(\mu, \nu) = \log(1 + \chi^2(\mu, \nu))$ is immediate from the definition. On the other hand, taking the limit as $q \to 1$ yields that $R_1(\mu, \nu) = \mathsf{KL}(\mu, \nu)$. Under the Poincaré inequality for μ, Vempala and Wibisono [2019, Theorem 3] show that for all $q \geqslant 2$:

$$R_q(\mu_t, \mu) \leqslant \max \left\{ R_q(\mu_0, \mu) - \frac{2t}{qC_{\mathsf{PI}}}, \exp\left(-\frac{2t}{qC_{\mathsf{PI}}} \right) R_q(\mu_0, \mu) \right\}. \tag{2.7}$$

That is, convergence in the q-Rényi divergence starts off in a linear regime, and then transitions to a exponential regime after $t = \frac{qC_{\mathsf{PI}}}{2} R_q(\mu_0, \mu)$ time. On the other hand, under the LSI assumption, Vempala and Wibisono [2019, Theorem 2] show that for all $q \geqslant 1$:

$$R_q(\mu_t, \mu) \leqslant \exp\left(-\frac{2t}{qC_{\mathsf{LSI}}} \right) R_q(\mu_0, \mu). \tag{2.8}$$

Since $R_1(\mu, \nu)$ recovers the KL-divergence, (2.8) generalizes Lemma 2.2. Comparing (2.8) to (2.7), we see that the stronger LSI assumption both increases the range of q for which the result applies, and removes the linear regime in (2.7). The proofs of (2.8) and (2.7) follow similar arguments as the proofs of Lemma 2.1 and Lemma 2.2.

3 A Lyapunov condition for the Poincaré inequality

In Section 2, we saw that convergence of the diffusion (1.1) was ensured via checking whether the stationary measure μ satisfied either the Poincaré or log-Sobolev inequality. As stated, it is not immediately clear how to check such functional inequalities. In this section, we follow the work of Bakry et al. [2008] to develop a Lyapunov condition which certifies the Poincaré inequality. Similar Lyapunov conditions also certify the LSI [Cattiaux et al., 2009]. We choose to focus on the Poincaré inequality because the proof is simpler.

Lyapunov analysis for deterministic dynamical systems works by finding a Lyapunov function V such that the time derivative of V along a trajectory is non-positive. In the diffusion setting, this principle still applies, but we need a new mechanism for taking a time derivative. For this purpose, we introduce the following diffusion operator:[2]

$$L := \Delta - \langle \nabla U, \nabla \rangle.$$

There are many paths of arriving at the operator L. The most direct is via Itô's lemma. In particular, for a smooth function f, Itô's lemma states that:

$$\begin{aligned} df(X_t) &= \left\{ -\langle \nabla f(X_t), \nabla U(X_t) \rangle + \mathrm{tr}(\nabla^2 f(X_t)) \right\} dt + \sqrt{2} \langle \nabla f(X_t), dB_t \rangle \\ &= (Lf)(X_t) dt + \sqrt{2} \langle \nabla f(X_t), dB_t \rangle. \end{aligned}$$

Thus, the operator L encodes the drift term for the Itô stochastic differential equation associated to $f(X_t)$.

We now have the necessary tools to describe a Lyapunov function for (1.1). To distinguish between Lyapunov functions for the deterministic dynamics $dX/dt = -\nabla U(X)$ (which we will consider in the next section) versus the diffusion (1.1), we will term Lyapunov functions for the latter as "stochastic Lyapunov functions".

Definition 3.1. *The function $W : \mathbb{R}^n \to \mathbb{R}$ is a (ρ, b, R)-*stochastic Lyapunov function *if the following conditions hold:*

1. $W \in C^2(\mathbb{R}^n)$ *and* $W \geqslant 1$.
2. $\rho > 0$, $R > 0$, *and* $b \geqslant 0$.
3. $LW \leqslant -\rho W + b \mathbf{1}_{B_2^n(R)}$.

We note that the third condition in Definition 3.1 is the continuous-time analog of the drift condition typically considered when studying ergodicty of discrete-time Markov chains [Meyn and Tweedie, 1993]. For a radius $R > 0$, let $\bar{\mu}_R$ be the measure supported on $B_2^n(R)$ with density:

$$d\bar{\mu}_R = \mathbf{1}_{B_2^n(R)} \frac{d\mu}{\int_{B_2^n(R)} d\mu}.$$

The following result states that if the diffusion (1.1) admits a stochastic Lyapunov function, and if the measure $\bar{\mu}_R$ satisfies a Poincaré inequality, then the full measure μ also satisfies a Poincaré inequality.

Lemma 3.2. *Let W be a (ρ, b, R)-stochastic Lyapunov function (cf. Definition 3.1). Suppose that $\bar{\mu}_R$ satisfies the Poincaré inequality with constant C_R. Then, μ satisfies the Poincaré inequality with constant $C_{\mathsf{PI}} = \frac{1 + C_R b}{\rho}$.*

Proof. This proof is from Bakry et al. [2008, Theorem 1.4]. Since ρW is positive, dividing the Lyapunov equation yields:

$$1 \leqslant -\frac{LW}{\rho W} + \frac{b}{\rho W} \mathbf{1}_{B_2^n(R)}. \tag{3.1}$$

[2]This operator is the infinitesimal generator of the Markov semigroup $(P_t)_{t \geqslant 0}$, where $(P_t f)(x) := \mathbb{E}[f(X_t) \mid X_0 = x]$ and $(X_t)_{t \geqslant 0}$ follows the process (1.1). See Bakry et al. [2014] for more details.

Fix a smooth function f. Since $\mathbb{E}_\mu f$ is the projection of $L^2(\mu)$ onto the span of constant functions, we have for any $c \in \mathbb{R}$:

$$\mathrm{Var}_\mu(f) \leqslant \mathbb{E}_\mu(f-c)^2.$$

We postpone the choice of c for the time being. Put $g := f - c$. Observe the following identity:

$$\begin{aligned}
\int g^2 \frac{LW}{\rho W} d\mu &= \int g^2 \frac{\Delta W}{\rho W} d\mu - \int g^2 \frac{\langle \nabla U, \nabla W\rangle}{\rho W} d\mu \\
&= -\int \left\langle \nabla \left(\frac{g^2 \mu}{\rho W}\right), \nabla W \right\rangle dx - \int g^2 \frac{\langle \nabla U, \nabla W\rangle}{\rho W} d\mu && \text{integration by parts} \\
&= -\int \left\langle \frac{g^2}{\rho W} \nabla \mu + \mu \nabla \left(\frac{g^2}{\rho W}\right), \nabla W \right\rangle dx - \int g^2 \frac{\langle \nabla U, \nabla W\rangle}{\rho W} d\mu \\
&= -\int \left\langle -\frac{g^2}{\rho W} \mu \nabla U + \mu \nabla \left(\frac{g^2}{\rho W}\right), \nabla W \right\rangle dx - \int g^2 \frac{\langle \nabla U, \nabla W\rangle}{\rho W} d\mu && \text{since } \nabla \mu = -\mu \nabla U \\
&= -\int \left\langle \nabla \left(\frac{g^2}{\rho W}\right), \nabla W \right\rangle d\mu.
\end{aligned}$$

Therefore by (3.1):

$$\begin{aligned}
\int g^2 d\mu &\leqslant -\int g^2 \frac{LW}{\rho W} d\mu + \int \frac{b}{\rho W} g^2 \mathbf{1}_{B_2^n(R)} d\mu \\
&\leqslant -\int g^2 \frac{LW}{\rho W} d\mu + \frac{b}{\rho} \int g^2 \mathbf{1}_{B_2^n(R)} d\mu && \text{since } W \geqslant 1 \\
&= \int \left\langle \nabla \left(\frac{g^2}{\rho W}\right), \nabla W \right\rangle d\mu + \frac{b}{\rho} \left(\int_{B_2^n(R)} d\mu\right) \int g^2 d\bar{\mu}_R. && (3.2)
\end{aligned}$$

We first handle the second term in (3.2). Choosing $c = \mathbb{E}_{\bar{\mu}_R} f$, by the Poincaré inequality for $\bar{\mu}_R$:

$$\int g^2 d\bar{\mu}_R = \mathrm{Var}_{\bar{\mu}_R}(f) \leqslant C_R \mathbb{E}_{\bar{\mu}_R} \|\nabla f\|^2 = C_R \int \|\nabla f\|^2 d\bar{\mu}_R \leqslant \frac{C_R}{\int_{B_2^n(R)} d\mu} \int \|\nabla f\|^2 d\mu.$$

Hence:

$$\frac{b}{\rho} \left(\int_{B_2^n(R)} d\mu\right) \int g^2 d\bar{\mu}_R \leqslant \frac{C_R b}{\rho} \int \|\nabla f\|^2 d\mu.$$

It remains to deal with the first term in (3.2). By completing the square:

$$\begin{aligned}
\int \left\langle \nabla \left(\frac{g^2}{\rho W}\right), \nabla W \right\rangle d\mu &= \frac{1}{\rho} \int \left\langle \frac{2g}{W} \nabla g - \frac{g^2}{W} \nabla W, \nabla W \right\rangle d\mu \\
&= \frac{1}{\rho} \int \left[\|\nabla g\|^2 - \|\nabla g\|^2 + \left\langle \frac{2g}{W} \nabla g - \frac{g^2}{W} \nabla W, \nabla W \right\rangle \right] d\mu \\
&= \frac{1}{\rho} \int \left[\|\nabla g\|^2 - \left\| \nabla g - \frac{g}{W} \nabla W \right\|^2 \right] d\mu \\
&\leqslant \frac{1}{\rho} \int \|\nabla g\|^2 d\mu
\end{aligned}$$

$$= \frac{1}{\rho}\int \|\nabla f\|^2 d\mu \qquad \text{since } g = f - c.$$

Therefore:

$$\mathrm{Var}_\mu(f) \leqslant \frac{1 + C_R b}{\rho}\int \|\nabla f\|^2 d\mu.$$

□

We now address the question of showing that the Poincaré inequality holds for the measure $\bar{\mu}_R$. First, we have the following Poincaré inequality for uniform measures on bounded sets. This result is standard in the theory of Sobolev spaces (see e.g. Evans [2010]).

Proposition 3.3. *Fix $R > 0$ and open set $\Omega \subset \mathbb{R}^n$. Suppose that $\sup_{x\in\Omega}\|x\|_\infty \leqslant R$. For any function $f \in C^1(\Omega)$, we have:*

$$\int_\Omega f^2 dx \leqslant 4R^2 \int_\Omega \|\nabla f\|^2 dx.$$

Proof. We extend f so that $f = 0$ on Ω^c. Now, we write $f = f(x_1, x')$, where $x_1 \in \mathbb{R}$ and $x' \in \mathbb{R}^{n-1}$. By the fundamental theorem of calculus, $f(x, x') = \int_{-R}^{x} \partial_{x_1} f(s, x') ds$. Now for a fixed x':

$$\begin{aligned} f^2(x, x') &= \left(\int_{-R}^{x} \partial_{x_1} f(s, x') ds\right)^2 \leqslant \left(\int_{-R}^{x} |\partial_{x_1} f(s, x')| ds\right)^2 \\ &\leqslant \left(\int_{-R}^{R} |\partial_{x_1} f(s, x')| ds\right)^2 \leqslant 2R \int_{-R}^{R} \|\nabla f(s, x')\|^2 ds, \end{aligned}$$

where the last inequality is Cauchy-Schwarz. Hence:

$$\int_\Omega f^2 dx \leqslant 2R \int_\Omega \int_{-R}^{R} \|\nabla f(s, x')\|^2 ds dx' \leqslant 4R^2 \int_\Omega \|\nabla f\|^2 dx.$$

□

For a radius $R > 0$, let us define

$$\mathrm{Osc}_R(U) := \sup_{x\in B_2^n(R)} U(x) - \inf_{x\in B_2^n(R)} U(x).$$

The following result gives a (crude) bound on the Poincaré constant for $\bar{\mu}_R$ in terms of $\mathrm{Osc}_R(U)$.

Proposition 3.4. *Fix an $R > 0$. The measure $\bar{\mu}_R$ satisfies the Poincaré inequality with $C_{\mathsf{PI}} = 4R^2 \exp(\mathrm{Osc}_R(U))$.*

Proof. Fix an $f \in C^1(B_2^n(R))$ and let $Z := \int_{B_2^n(R)} \exp(-U) dx$. We have:

$$\begin{aligned} \int f^2 d\bar{\mu}_R &= \int_{B_2^n(R)} f^2 \frac{\exp(-U)}{Z} dx \\ &\leqslant Z^{-1} \sup_{x\in B_2^n(R)} \exp(-U(x)) \int_{B_2^n(R)} f^2 dx \end{aligned}$$

$$
\begin{aligned}
&\leqslant 4R^2 Z^{-1} \sup_{x\in B_2^n(R)} \exp(-U(x)) \int_{B_2^n(R)} \|\nabla f\|^2 dx && \text{using Proposition 3.3} \\
&= 4R^2 \sup_{x\in B_2^n(R)} \exp(-U(x)) \int_{B_2^n(R)} \|\nabla f\|^2 \frac{\exp(-U)}{Z\exp(-U)} dx \\
&\leqslant 4R^2 \exp(\mathrm{Osc}_R(U)) \int \|\nabla f\|^2 d\bar{\mu}_R.
\end{aligned}
$$

□

4 From deterministic to stochastic Lyapunov functions

In Section 3, we established the notion of a stochastic Lyapunov function, which certified the Poincaré inequality for μ. The purpose of this section is to relate Lyapunov functions for the deterministic system $dX/dt = -\nabla U(X)$ to their associated stochastic Lyapunov functions.

The following definition gives us a class of Lyapunov functions for $dX/dt = -\nabla U(X)$.

Definition 4.1. *Fix positive constants ρ, μ, L and $x_\star \in \mathbb{R}^n$. The function $V : \mathbb{R}^n \to \mathbb{R}$ is a $(\rho, \mu, L, x_\star)$-*Lyapunov function *if $V \in C^2(\mathbb{R}^n)$, and for all $x \in \mathbb{R}^n$:*

1. $V(x) \geqslant \mu\|x - x_\star\|^2$,
2. $\langle -\nabla U(x), \nabla V(x)\rangle \leqslant -\rho V(x)$,
3. $\nabla^2 V \preccurlyeq LI$.

The first two conditions are standard for exponential convergence. The minorization condition $V(x) \geqslant \mu\|x - x_\star\|^2$ can be replaced with a more general condition $V(x) \geqslant \mu(\|x - x_\star\|)$ where μ is a class-$\mathcal{K}$ function, but we will not do this for simplicity. The last condition is not typical of standard Lyapunov analysis, and is equivalent to stating that V has L-Lipschitz gradients. This condition is necessary to ensure that the Laplacian term in the operator L is bounded. One can relax this third condition to allow for mild growth of the Hessian: $\nabla^2 V \preccurlyeq \max\{L, \|x - x_\star\|^\alpha\} I$ for $0 < \alpha < 2$ (or more generally if condition one is replaced by a general class-$\mathcal{K}$ function, then the Hessian growth must be dominated by the function that minorizes V). Again, we do not do this for simplicity.

The next result shows that setting $W = 1 + V$ generates a valid stochastic Lyapunov function.

Proposition 4.2. *Suppose that V is a $(\rho, \mu, L, x_\star)$-Lyapunov function (cf. Definition 4.1). Then the function $W := 1 + V$ is a $(\rho/2, \rho + Ln, R)$-stochastic Lyapunov function (cf. Definition 3.1) for*

$$
R := \|x_\star\| + \sqrt{\frac{2}{\mu}\left(1 + \frac{Ln}{\rho}\right)}.
$$

Proof. Recall that $L = \Delta - \langle \nabla U, \nabla\rangle$. Hence:

$$
\begin{aligned}
LW &= \Delta W - \langle \nabla U, \nabla W\rangle \\
&= \Delta V - \langle \nabla U, \nabla V\rangle && W = 1 + V \\
&\leqslant -\rho V + \Delta V && -\langle \nabla U, \nabla V\rangle \leqslant -\rho V \\
&\leqslant -\rho W + \rho + Ln && \nabla^2 V \preccurlyeq LI
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{\rho}{2}W - \frac{\rho}{2}W + \rho + Ln \\
&\leqslant -\frac{\rho}{2}W - \frac{\mu\rho}{2}\|x - x_\star\|^2 + \rho + Ln && W(x) \geqslant V(x) \geqslant \mu\|x - x_\star\|^2 \\
&\leqslant -\frac{\rho}{2}W + (\rho + Ln)\mathbf{1}\left\{\|x - x_\star\| \leqslant \sqrt{\frac{2}{\mu}\left(1 + \frac{Ln}{\rho}\right)}\right\} \\
&\leqslant -\frac{\rho}{2}W + (\rho + Ln)\mathbf{1}\left\{\|x\| \leqslant \|x_\star\| + \sqrt{\frac{2}{\mu}\left(1 + \frac{Ln}{\rho}\right)}\right\}.
\end{aligned}
$$

□

Our final result combines all the elements together. It gives a rate of convergence for the process (1.1), in terms of Lyapunov stability analysis for the associated deterministic dynamics $dX/dt = -\nabla U(X)$.

Theorem 4.3. *Suppose that V is a $(\rho, \mu, L, x_\star)$-Lyapunov function (cf. Definition 4.1). Then μ satisfies the Poincaré inequality with constant:*

$$
C_{\mathsf{PI}} = \frac{2}{\rho} + 8\left(1 + \frac{Ln}{\rho}\right)R^2\exp(\mathrm{Osc}_R(U)), \;\; R := \|x_\star\| + \sqrt{\frac{2}{\mu}\left(1 + \frac{Ln}{\rho}\right)}.
$$

For any initial measure μ_0, we have:

$$
\chi^2(\mu_t, \mu) \leqslant \exp\left(-\frac{2t}{C_{\mathsf{PI}}}\right)\chi^2(\mu_0, \mu).
$$

Proof. First, Proposition 4.2 yields that $W = 1 + V$ is a $(\rho/2, \rho + Ln, R)$-stochastic Lyapunov function. Next, by Proposition 3.4, the measure $\bar{\mu}_R$ satisfies the Poincaré inequality with $C_R = 4R^2\exp(\mathrm{Osc}_R(U))$. Lemma 3.2 then establishes that μ satisfies the Poincaré inequality with constant:

$$
\frac{2}{\rho} + 8\left(1 + \frac{Ln}{\rho}\right)R^2\exp(\mathrm{Osc}_R(U)).
$$

The convergence result follows by Lemma 2.1. □

4.1 Log-concave distributions

An important setting for Langevin diffusion is when the potential function U is convex, or equivalently, when the measure μ is log-concave. In this setting, we can easily construct a Lyapunov function.

Proposition 4.4. *Fix constants α, β satisfying $0 < \alpha \leqslant \beta < \infty$. Suppose that $U \in C^2(\mathbb{R}^n)$ satisfies $\alpha I \preccurlyeq \nabla^2 U(x) \preccurlyeq \beta I$ for all $x \in \mathbb{R}^n$. Put $x_\star = \mathrm{argmin}_{x\in\mathbb{R}^n} U(x)$. Then, the function $V(x) := \frac{1}{2}\|x - x_\star\|^2$ is a $(2\alpha\beta/(\alpha+\beta), 1/2, 1, x_\star)$-Lyapunov function (cf. Definition 4.1).*

Proof. A standard property [see e.g. Bubeck, 2015, Lemma 3.11] of α-strongly convex and β-smooth functions is that for all $x, y \in \mathbb{R}^n$:

$$
\langle \nabla U(x) - \nabla U(y), x - y\rangle \geqslant \frac{\alpha\beta}{\alpha+\beta}\|x - y\|^2 + \frac{1}{\alpha+\beta}\|\nabla U(x) - \nabla U(y)\|^2.
$$

Setting $y = x_\star$, we have:

$$\langle \nabla U(x), \nabla V(x) \rangle = \langle \nabla U(x), x - x_\star \rangle \geqslant \frac{\alpha\beta}{\alpha+\beta}\|x - x_\star\|^2 = \frac{2\alpha\beta}{\alpha+\beta}V(x).$$

□

While Proposition 4.4 is simple, it yields a very suboptimal Poincaré inequality constant (and hence suboptimal constants in the convergence rate). It turns out that if U is α-strongly convex, then μ can be shown to satisfies the (stronger) log-Sobolev inequality with constant $C_{\mathsf{LSI}} = 1/\alpha$. This condition is known as the Bakry-Émery criterion. The proof of this result is beyond the scope of this paper.

Theorem 4.5 (Bakry et al. [2014, Corollary 5.7.2])**.** *Fix $\alpha > 0$. Suppose $U \in C^2(\mathbb{R}^n)$ satisfies the condition $\nabla^2 U(x) \succcurlyeq \alpha I$ for all $x \in \mathbb{R}^n$. Then, $\mu = \exp(-U)$ satisfies the log-Sobolev inequality with $C_{\mathsf{LSI}} = 1/\alpha$.*

Acknowledgements

I thank Nicholas M. Boffi for both sparking my interest in this subject, and also patiently explaining many of the concepts in this manuscript to me.

References

Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin. A simple proof of the poincaré inequality for a large class of probability measures. *Electronic Communications in Probability*, 13:60–66, 2008.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2014.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–357, 2015.

Patrick Cattiaux, Arnaud Guillin, Feng-Yu Wang, and Liming Wu. Lyapunov conditions for super poincaré inequalities. *Journal of Functional Analysis*, 256(6):1821–1841, 2009.

Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.

Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

A Omitted proofs

Proposition A.1. *Suppose that μ satisfies the log-Sobolev inequality with constant C_{LSI}. Then, μ satisfies the Poincaré inequality with constant $C_{\mathsf{PI}} = C_{\mathsf{LSI}}$.*

Proof. This proof is from Bakry et al. [2014, Proposition 5.1.3]. Define $h(x) := x \log x$. Let g be a function with $\mathbb{E}_\mu g = 0$. Fix $\varepsilon > 0$, and set $\bar{f} = 1 + \varepsilon g$. We have:

$$\mathrm{Ent}_\mu(\bar{f}^2) = \mathbb{E}_\mu h((1+\varepsilon g)^2) - h(1 + \varepsilon^2 \mathbb{E}_\mu g^2) =: h_1(\varepsilon) - h_2(\varepsilon).$$

Computing the first and second derivative of h_1 and h_2:

$$\begin{aligned}
h_1'(\varepsilon) &= 2\mathbb{E}_\mu (1 + \log((1+\varepsilon g)^2))(1+\varepsilon g) g, \\
h_1''(\varepsilon) &= \mathbb{E}_\mu \left[(1 + \log((1+\varepsilon g)^2)) 2g^2 + 4g^2 \right], \\
h_2'(\varepsilon) &= (1 + \log(1 + \varepsilon^2 \mathbb{E}_\mu g^2)) 2\varepsilon \mathbb{E}_\mu g^2, \\
h_2''(\varepsilon) &= (1 + \log(1 + \varepsilon^2 \mathbb{E}_\mu g^2)) 2\mathbb{E}_\mu g^2 + \frac{4\varepsilon^2 (\mathbb{E}_\mu g^2)^2}{1 + \varepsilon^2 \mathbb{E}_\mu g^2}.
\end{aligned}$$

Taking a Taylor expansion of $\mathrm{Ent}_\mu(\bar{f}^2)$:

$$\begin{aligned}
\mathrm{Ent}_\mu(\bar{f}^2) &= h_1(0) + h_1'(0)\varepsilon + \frac{h_1''(0)}{2}\varepsilon^2 - h_2(0) - h_2'(0)\varepsilon - \frac{h_2''(0)}{2}\varepsilon^2 + o(\varepsilon^2) \\
&= 2\mathbb{E}_\mu g^2 \varepsilon^2 + o(\varepsilon^2).
\end{aligned}$$

By the log-Sobolev inequality:

$$\mathrm{Ent}_\mu(\bar{f}^2) \leqslant 2C_{\mathsf{LSI}} \mathbb{E}_\mu \|\nabla \bar{f}\|^2 = \varepsilon^2 2C_{\mathsf{LSI}} \mathbb{E}_\mu \|\nabla g\|^2.$$

Hence taking the limit as $\varepsilon \to 0$:

$$\mathbb{E}_\mu g^2 \leqslant C_{\mathsf{LSI}} \mathbb{E}_\mu \|\nabla g\|^2.$$

Now for any f, set $g = f - \mathbb{E}_\mu f$, from which the result follows. $\square$