Université PSL (Paris Sciences & Lettres)
Master *Mathématiques théoriques et appliquées*

Logarithmic Sobolev Inequalities Essentials

Rough lecture notes

Djalil Chafaï & Joseph Lehec

Winter 2017, Université Paris-Dauphine, Paris
Revised Spring 2017, Universidad de Chile, Santiago de Chile
Revised Winter 2023, École normale supérieure, Paris

This course is a modern overview on logarithmic Sobolev inequalities, from the probabilistic side. These inequalities have been the subject of intense activity in the recent decades in relation with the analysis and geometry of Markov processes and diffusion evolution equations. This course is designed to be accessible to a wide audience with a first year of master level in mathematics. It is divided into seven lectures of roughly three hours.

Short bibliography:

- Bakry, Gentil, and Ledoux – *Analysis and Geometry of Markov Diffusion Operators*
- Villani – *Optimal transport, old and new*
- Royer – *An Initiation to Logarithmic Sobolev inequalities*
- Ané et al – *Sur les inégalités de Sobolev logarithmiques*

Contents

Chapter 1

Introduction

The first chapters focus on a Gaussian model formed with the product space $\mathbb{R}^n$ equipped with the Gaussian probability measure $\gamma_n = \gamma_1^{\otimes n}$. It allows explicit computations. It appears asympotically in other models due to the central limit phenomenon[1], in particular from spheres and from cubes, both equipped with the uniform measure. In a way γ_n plays the role of a uniform measure on $\mathbb{R}^n$.

The standard Gaussian measure γ_n on $\mathbb{R}^n$ has expectation 0, covariance the identity matrix I_n, and density with respect to the Lebesgue measure given by

$$x \mapsto (2\pi)^{-n/2}\,\mathrm{e}^{-|x|^2/2}$$

where $|x| = \left(\sum_{i\leq n} x_i^2\right)^{1/2}$ denotes the Euclidean norm.

Let $\mathcal{U}_n$ be the uniform distribution on the unit sphere $\{s \in \mathbb{R}^n : |s| = 1\}$.

Theorem 1.1 (Polar factorization of spheres)**.** *If $X = (X_1, \ldots, X_n) \sim \gamma_n$ then*

$$|X| = \sqrt{X_1^2 + \cdots + X_n^2} \quad \text{and} \quad \frac{X}{|X|}$$

are independent, with $|X|^2 \sim \chi^2(n)$ and $X/|X| \sim \mathcal{U}_n$. Conversely if R and U are independent with $R^2 \sim \chi^2(n)$ and $U \sim \mathcal{U}_n$, then $RU \sim \gamma_n$.

Proof. Follows from $\mathrm{e}^{-|x|^2/2}dx = \mathrm{e}^{-r^2/2}r^{n-1}drdu$ where $x = ru$. □

The following result, known as the Borel or Poincaré observation, states that one can see the standard Gaussian as a the projection of the uniform distribution on high dimensional spheres with radius equal to the square root of the dimension.

Theorem 1.2 (Central Limit Theorem for spheres)**.** *If $U_n \sim \mathcal{U}_n$ for any $n \geq 1$, then for any fixed $k \geq 1$, $\mathrm{proj}(\sqrt{n}U_n, \mathbb{R}^k) \to \gamma_k$ in law as $n \to \infty$.*

Proof. Let $(X_n)_{n\geq 1}$ be independent and identically distributed random variables with law γ_1. By the preceding theorem and the strong law of large numbers,

$$\mathrm{proj}(\sqrt{n}U_n, \mathbb{R}^k) \overset{d}{=} \sqrt{n}\frac{(X_1, \ldots, X_k)}{\sqrt{X_1^2 + \cdots + X_n^2}} \xrightarrow[n\to\infty]{a.s.} (X_1, \ldots, X_k) \sim \gamma_k.$$

□

[1] If $X_1, X_2, \ldots$ are i.i.d. real random variables with zero mean and unit variance then $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges in law as $n \to \infty$ to the standard Gaussian distribution.

Definition 1.3 (Variance and entropy)**.** If f is a square integrable function with respect to γ_n, then its *variance* is

$$\mathrm{Var}_{\gamma_n}(f) = \int_{\mathbb{R}^n} f^2 \,\mathrm{d}\gamma_n - \left(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma_n\right)^2 = \mathrm{Var}(f(X)) \quad \text{where } X \sim \gamma_n.$$

If f is a non negative function, integrable with respect to γ_n, then its *entropy* is

$$\mathrm{Ent}_{\gamma_n}(f) = \int_{\mathbb{R}^n} f \log f \,\mathrm{d}\gamma_n - \left(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma_n\right) \log\left(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma_n\right).$$

The function $f \log f$ may not be integrable, but since the function $x \log x$ is bounded from below, the integral of $f \log f$ always makes sense in $\mathbb{R} \cup \{+\infty\}$.

Remark (ϕ entropies)**.** *One can define a more general object: Given a convex function ϕ on an interval $I \subset \mathbb{R}$ the ϕ–entropy of $f : \mathbb{R} \to I$ is*

$$\mathbb{E}^{\phi}_{\gamma_n}(f) = \int_{\mathbb{R}^n} \phi(f) \,\mathrm{d}\gamma_n - \phi\left(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma_n\right).$$

We recover the variance for $\phi(x) = x^2$ and $I = \mathbb{R}$ and the entropy for $\phi(x) = x \log(x)$ and $I = \mathbb{R}_+$. By Jensen's inequality, a ϕ–entropy is always non negative, and if the function ϕ is strictly convex, which is indeed the case for x^2 and $x \log(x)$, then the ϕ–entropy only vanishes on constant functions.

The Poincaré inequality for the Gaussian measure states as follows:

$$\mathrm{Var}_{\gamma_n}(f) \le \int_{\mathbb{R}^n} |\nabla f|^2 \,\mathrm{d}\gamma_n,$$

while the logarithmic Sobolev inequality reads

$$\mathrm{Ent}_{\gamma_n}(f) \le \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \,\mathrm{d}\gamma_n.$$

We will explain in the sequel why these are interesting inequalities and how they are related to the Ornstein–Uhlenbeck process and its numerous properties.

1.1 Generalities on Markov processes

In this section, we give a very brief recap on Markov processes. We give the main definitions alongside with a couple of very basic properties.

Let E be a set and $\mathcal{B}$ be a σ–field on E. A process (X_t) taking values in E is said to be Markovian if for every $t \ge 0$, the conditional law of $(X_s)_{s \ge t}$ given $\mathcal{F}_t = \sigma(X_{s \le t})$ coincides with the law of $(X_s)_{s \ge t}$ given X_t and with the law of $(X_s)_{s \ge 0}$ given X_0. We associate a semigroup to the process (X_t): Let

$$P_t f(x) = \mathbb{E}_x[f(X_t)]$$

for every bounded and measurable function f, for every $t \ge 0$, for every $x \in E$, and where $\mathbb{E}_x$ denotes expectation given $X_0 = x$.

Proposition 1.4 (Markov semigroup)**.** *For every $s, t \ge 0$, we have*

(i) $P_{s+t} = P_s \circ P_t$;

(ii) If $f \geq 0$ *then* $P_t f \geq 0$*;*

(iii) $P_t \mathbf{1} = \mathbf{1}$*.*

We say that (P_t) *is a* Markov semigroup*.*

Proof. Since (X_t) is a Markov process and by definition of P_t

$$\mathbb{E}[f(X_{s+t}) \mid X_s] = P_t f(X_s)$$

Taking expectation again we get

$$P_{s+t} f(x) = P_s(P_t f)(x),$$

which is the first property. The other two properties are obvious from the definition. □

The semigroup (P_t) also acts on measures through the duality

$$\int_E f \, d(\mu P_t) = \int_E P_t f \mathrm{d}\mu.$$

If μ is a probability measure, this has a probabilistic interpretation: μP_t is the law of X_t under $\mathbb{P}_\mu$, in other words the law of X_t when X_0 has law μ.

Definition 1.5 (Stationary measure)**.** A measure μ is called stationary if $\mu P_t = \mu$ for all t, in other words if

$$\int_E P_t f \, \mathrm{d}\mu = \int_E f \, \mathrm{d}\mu, \quad \forall t \geq 0$$

for every bounded and measurable f.

Remark. *If* μ *is a probability measure, this can be reformulated as*

$$X_0 \sim \mu \quad \Rightarrow \quad X_t \sim \mu, \ \forall t \geq 0.$$

Stationary measures have the following fundamental property.

Lemma 1.6 (Contractivity)**.** *If* μ *is stationary then* P_t *extends to a continuous operator on* $\mathbb{L}^p(\mu)$ *for any* $p \in [1, \infty]$*. Moreover* P_t *is a contraction:*

$$\|P_t f\|_p \leq \|f\|_p, \quad \forall p \in [1, \infty], \quad \forall f \in \mathbb{L}^p(\mu).$$

Proof. Let f be bounded and in $\mathbb{L}^p(\mu)$. By Jensen's inequality we have $|P_t f|^p \leq P_t(|f|^p)$, pointwise. Integrating and using stationarity we get

$$\int_E |P_t f|^p \, \mathrm{d}\mu \leq \int_E P_t(|f|^p) \, \mathrm{d}\mu = \int_E |f|^p \, \mathrm{d}\mu.$$

Since bounded functions are dense in $\mathbb{L}^p(\mu)$ this is the result. □

It is pretty clear from the semigroup property, that (P_t) is completely determined by its behavior when t tends to 0. For this reason, it is natural to try and differentiate $P_t f$ at $t = 0$. For functions f such that

$$\lim_{t \to 0} \frac{P_t f - f}{t} \tag{1.1}$$

exists (let us not specify in what sense at this stage) we let Lf be this limit. The operator L is called the *generator* of the semigroup P_t. Then using the semigroup property and assuming that P_t is continuous for whatever topology was considered at (1.1) we get

$$LP_t f = \lim_{s\to 0} \frac{P_s(P_t f) - P_t f}{s} = P_t \left(\lim_{s\to 0} \frac{P_s f - f}{s} \right) = P_t(Lf).$$

So P_t and L commute.

Let us give a last definition before moving on to the particular case of the Ornstein–Uhlenbeck semigroup.

Definition 1.7 (Reversible measure)**.** A measure μ is called reversible if

$$\int_E (P_t f) g \,\mathrm{d}\mu = \int_E f(P_t g)\,\mathrm{d}\mu$$

for every bounded and measurable f, g.

Remarks. *Reversibility is stronger than stationarity (just take $g = 1$). Reversibility also has a probabilistic interpretation, a probability measure μ is reversible if and only if*

$$X_0 \sim \mu \quad \Rightarrow \quad (X_0, X_t) \sim (X_t, X_0), \ \forall t \geq 0.$$

At the level of the generator reversibility reads

$$\int_E (Lf) g \,\mathrm{d}\mu = \int_E f(Lg)\,\mathrm{d}\mu.$$

In other words L is a symmetric operator on $\mathbb{L}^2(\mu)$.

1.2 Ornstein–Uhlenbeck semigroup

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)$ carrying a standard n–dimensional Brownian motion (B_t). We consider the following stochastic differential equation

$$dX_t = \sqrt{2}\,dB_t - X_t\,\mathrm{d}t. \tag{1.2}$$

Recall that a process (X_t) is a solution to (1.2) if (X_t) is adapted to the filtration $(\mathcal{F}_t)$ and if

$$X_t = \sqrt{2}B_t - \int_0^t X_s\,\mathrm{d}s, \quad \forall t \geq 0$$

almost surely. Note that this imply implicitly that the integral above should be well defined, so we should have

$$\int_0^t |X_s|\,\mathrm{d}s < +\infty, \quad \forall t \geq 0,$$

almost surely. Actually (1.2) can be solved explicitly. Indeed it implies that

$$d(\mathrm{e}^t X_t) = \mathrm{e}^t dX_t + \mathrm{e}^t X_t \mathrm{d}t = \sqrt{2}\mathrm{e}^t\,dB_t.$$

Hence

$$X_t = \mathrm{e}^{-t} X_0 + \sqrt{2} \int_0^t \mathrm{e}^{s-t}\,dB_s. \tag{1.3}$$

Conversely, if (X_t) is defined by (1.3) then it clearly satisfies all the previous requirements. We claim that (X_t) is a Markov process. Indeed, an easy computation shows that

$$X_{t+s} = \mathrm{e}^{-t}\, X_s + \sqrt{2} \int_0^t \mathrm{e}^{u-t}\, d\widetilde{B}_u$$

where $\widetilde{B}_u = B_{s+u} - B_s$. Since $\widetilde{B}$ is a Brownian motion independent of $\mathcal{F}_s$ we obtain that the conditional law of $(X_{t+s})_{t\geq 0}$ given $\mathcal{F}_s$ coincides with that of an independent OU process $(\widetilde{X}_t)$ initiated from X_s.

Let (P_t) be the associated semigroup. Note that $\sqrt{2}\int_0^t \mathrm{e}^{s-t}\, dB_s$ is a Gaussian vector centered at 0 and having covariance matrix

$$2\left(\int_0^t \mathrm{e}^{2(s-t)}\, \mathrm{d}s\right) I_n = (1-\mathrm{e}^{2t})\, I_n.$$

Therefore if $X_0 = x$ then $X_t \sim \mathcal{N}(\sqrt{\rho}\, x, (1-\rho)\, I_n)$ where $\rho = \mathrm{e}^{-2t}$. We thus have the following expression for $P_t f$, called the Mehler formula.

Lemma 1.8 (Mehler formula)**.** *For every test function f and every $x \in \mathbb{R}^n$*

$$\begin{aligned} P_t f(x) &= \int_{\mathbb{R}^n} f\left(\sqrt{\rho}\, x + \sqrt{1-\rho}\, y\right) \gamma_n(dy) \\ &= f * g_{1-\rho}(\sqrt{\rho}\, x), \end{aligned}$$

where $\rho = \mathrm{e}^{-2t}$, and $g_{1-\rho}$ is the density of the $\mathcal{N}(0,(1-\rho)I_n)$ law.

Remark. *The semigroup property $P_{s+t} = P_s \circ P_t$ is easily retrieved using convolution properties of the Gaussian density.*

Lemma 1.9. *The standard Gaussian measure is reversible for (P_t).*

Proof. Let $\rho = \mathrm{e}^{-2t}$. From (1.3) we see that if X_0 is a standard Gaussian vector then the couple (X_0, X_t) is Gaussian on $\mathbb{R}^{2n}$ with expectation 0 and covariance matrix

$$\begin{pmatrix} I_n & \sqrt{\rho}\, I_n \\ \sqrt{\rho}\, I_n & I_n \end{pmatrix}.$$

In particular $(X_0, X_t) = (X_t, X_0)$, in law. □

Lemma 1.10. *For any initial distribution μ, we have $X_t \to \gamma_n$, in law.*

Proof. Let G be a standard Gaussian vector independent of X_0. We have seen that

$$X_t = \sqrt{\rho}\, X_0 + \sqrt{1-\rho}\, G, \quad \text{in law}.$$

If $t \to +\infty$ then $\rho \to 1$ and

$$\sqrt{\rho}\, X_0 + \sqrt{1-\rho}\, G \to G$$

almost surely. Hence the result. □

Corollary 1.11. *γ_n is the only stationary distribution of the process.*

Proof. Let μ be a probability distribution satisfying $\mu P_t = \mu$ for all t. Taking the limit as $t \to \infty$ yields $\mu = \gamma_n$. □

Now we compute the generator of the Ornstein–Uhlenbeck semigroup. Observe first that the path (X_t) of the Ornstein–Uhlenbeck process is almost surely continuous. For every t we have $\lim_{s\to t} X_s = X_t$ almost surely, hence in law. Therefore for every fixed x and continuous and bounded function f we have $\lim_{s\to t} P_s f(x) = P_t f(x)$. In other words the map $t \mapsto P_t f(x)$ is continuous. We now investigate its differentiability. Let $\mathcal{C}^2$ be the space of twice continuously differentiable functions which are bounded with bounded partial derivatives of order 1 and 2.

Lemma 1.12 (Infinitesimal generator)**.** *Let $f \in \mathcal{C}^2$ and let*

$$Lf(x) = \Delta f(x) - \langle \nabla f(x), x\rangle.$$

Then for every $x \in \mathbb{R}^n$ we have

$$\lim_{t\to 0} \frac{P_t f(x) - f(x)}{t} = Lf(x).$$

Actually for fixed x in $\mathbb{R}^n$ the map $t \mapsto P_t f(x)$ is continuously differentiable and

$$\partial_t P_t f(x) = P_t(Lf)(x).$$

Lastly, P_t preserves the class $\mathcal{C}^2$ and commutes with L, namely $L(P_t f) = P_t(Lf)$.

Proof. By definition of (X_t) and applying Itô' formula we have

$$\begin{aligned} f(X_t) - f(X_0) &= \int_0^t \langle \nabla f(X_s), dX_s\rangle + \int_0^t \Delta f(X_s)\mathrm{d}s \\ &= \int_0^t \langle \nabla f(X_s), \sqrt{2}dB_s - X_s\mathrm{d}s\rangle + \Delta f(X_s)\mathrm{d}s \qquad (1.4) \\ &= \sqrt{2}\int_0^t \langle \nabla f(X_s), dB_s\rangle + \int_0^t Lf(X_s)\mathrm{d}s. \end{aligned}$$

The hypothesis made on f imply in particular that $(|\nabla f(X_t)|)$ is bounded. So the stochastic integral of the previous equation is a martingale. In particular it has expectation 0. So taking expectation with respect to $\mathbb{P}_x$ in (1.4) yields the "Duhamel formula"

$$P_t f(x) - f(x) = \int_0^t P_s(Lf)(x)\mathrm{d}s.$$

Since Lf is continuous and bounded, $s \to P_s(Lf)(x)$ is continuous and we get the first part of the Lemma. The fact that P_t preserves $\mathcal{C}^2$ is clear from Mehler's formula, and we have explained why P_t and L commute in the previous section. □

Remark. *The formula for the generator L can also be derived directly from the Mehler formula using a Taylor formula or the fact that g_t solves the heat equation (we speak about the heat kernel): $\partial_t g_t = \frac{1}{2}\Delta g_t$. We used the SDE* (1.2) *and Itô's formula instead because this proof extends to processes whose semigroups do not have an explicit expression.*

Observe that if f and g belong to $\mathcal{C}^2$ then so does their product fg. Then an easy computation shows that

$$L(fg) = (Lf)g + f(Lg) + 2\langle \nabla f, \nabla g\rangle. \qquad (1.5)$$

Lemma 1.13 (Integration by parts formula)**.** *For f and g in $\mathcal{C}^2$ we have the following integration by parts formula*

$$\int_{\mathbb{R}^n} (Lf)g\,\mathrm{d}\gamma_n = -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle\,\mathrm{d}\gamma_n.$$

Proof. By dominated convergence, for $f, g \in \mathcal{C}^2$ we can differentiate at $t = 0$ the equality

$$\int_{\mathbb{R}^n} (P_t f) g \,\mathrm{d}\gamma_n = \int_{\mathbb{R}^n} f(P_t g) \,\mathrm{d}\gamma_n.$$

and get

$$\int_{\mathbb{R}^n} (Lf) g \,\mathrm{d}\gamma_n = \int_{\mathbb{R}^n} f(Lg) \,\mathrm{d}\gamma_n.$$

In the same way, we have

$$\int_{\mathbb{R}^n} L(fg) \,\mathrm{d}\gamma_n = 0.$$

Plugging in (1.5) and using the previous equality, we get the result. □

For a general Markov process having generator L the operator

$$\Gamma(f,g) = \frac{1}{2}\left(L(fg) - (Lf)g - f(Lg)\right)$$

is a fundamental object called *carré du champ* (even in English). The above proof shows that if there is a reversible measure μ we will have

$$-\int (Lf) g \,\mathrm{d}\mu = \int \Gamma(f,g) \,\mathrm{d}\mu$$

for every f, g in the appropriate space. This quantity is called the Dirichlet form (this is general notion of quadratic forms for unbounded operators).

1.3 Poincaré and logarithmic Sobolev inequalities

Lemma 1.14 (Semigroup expression of entropies)**.** *Let $\phi : I \to \mathbb{R}$ be convex and $\mathcal{C}^2$ on an interval $I \subset \mathbb{R}$ and $f\colon \mathbb{R}^n \to I$ be $\mathcal{C}^2$. Then*

$$\mathbb{E}^{\phi}_{\gamma_n}(f) = \int_0^{+\infty} \int_{\mathbb{R}^n} \phi''(P_t f) |\nabla P_t f|^2 \,\mathrm{d}\gamma_n \mathrm{d}t. \tag{1.6}$$

Proof. Let us give a formal proof first. Let us define, for all $t \geq 0$,

$$\alpha(t) = \int_{\mathbb{R}^n} \phi(P_t f) \,\mathrm{d}\gamma_n.$$

We have $\alpha(0) = \displaystyle\int \phi(f) \,\mathrm{d}\gamma$ and

$$\lim_{t\to+\infty} \alpha(t) = \int_{\mathbb{R}^n} \lim_{t\to+\infty} \phi(P_t f) \,\mathrm{d}\gamma_n = \phi\left(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma_n\right).$$

Therefore

$$\mathbb{E}^{\phi}_{\gamma_n}(f) = \alpha(0) - \alpha(+\infty).$$

Now using chain rule and integration by parts we get

$$\begin{aligned} \alpha'(t) = \int_{\mathbb{R}^n} \partial_t \phi(P_t f) &= \int_{\mathbb{R}^n} \phi'(P_t f)(LP_t f) \,\mathrm{d}\gamma_n \\ &= -\int_{\mathbb{R}^n} \langle \nabla \phi'(P_t f), \nabla P_t f \rangle \,\mathrm{d}\gamma_n \\ &= -\int_{\mathbb{R}^n} \phi''(P_t f) |\nabla P_t f|^2 \,\mathrm{d}\gamma_n, \end{aligned}$$

hence the result. Now we let the reader check that every step of this argument is valid if f belongs to $\mathcal{C}^2$ and takes values in some interval I on which ϕ is twice continuously differentiable with bounded derivatives. □

Lemma 1.15. *If f is smooth with bounded derivative then*

$$\nabla P_t f(x) = \mathrm{e}^{-t} P_t(\nabla f)(x)$$

Remark. $P_t(\nabla f)$ *is defined by extending* P_t *to* $\mathbb{R}^n$*–valued function coordinate wise.*

Proof. This is a clear consequence of Mehler's formula. □

Corollary 1.16. *Under the same hypothesis, the following inequalities hold true*

$$|\nabla P_t f|^2 \leq \mathrm{e}^{-2t} P_t(|\nabla f|^2) \tag{1.7}$$

$$\frac{|\nabla P_t f|^2}{P_t f} \leq \mathrm{e}^{-2t} P_t\left(\frac{|\nabla f|^2}{f}\right). \tag{1.8}$$

Proof. By the previous lemma and Cauchy–Schwarz we get

$$|\nabla P_t f|^2 = \mathrm{e}^{-2t} |P_t(\nabla f)|^2 \leq \mathrm{e}^{-2t} P_t(|\nabla f|^2)$$

which is the first inequality. Using Cauchy–Scwharz in a different way:

$$|P_t(\nabla f)|^2 = \left| P_t\left(\frac{\nabla f}{\sqrt{f}} \sqrt{f}\right)\right|^2 \leq P_t\left(\frac{|\nabla f|^2}{f}\right) P_t(f).$$

yields the second one. □

Theorem 1.17 (Poincaré inequality)**.** *For every function f whose gradient belongs to* $\mathbb{L}^2(\gamma_n)$ *we have*

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 \,\mathrm{d}\gamma_n.$$

Exercise (Optimality)**.** *Check on affine functions $f(x) = \langle x, u\rangle$ that the inequality is sharp.*

Proof. Clearly, if f belongs to $\mathcal{C}^2$ and $\phi(x) = x^2$, equality (1.6) applies and we get

$$\mathrm{Var}_{\gamma_n}(f) = 2\int_0^{+\infty} \int_{\mathbb{R}^n} |\nabla P_t f|^2 \,\mathrm{d}\gamma_n.$$

Moreover, by (1.7) and stationarity

$$\int_{\mathbb{R}^n} |\nabla P_t f|^2 \,\mathrm{d}\gamma_n \leq \mathrm{e}^{-2t} \int_{\mathbb{R}^n} P_t(|\nabla f|^2) \,\mathrm{d}\gamma_n = \mathrm{e}^{-2t} \int_{\mathbb{R}^n} |\nabla f|^2 \,\mathrm{d}\gamma_n.$$

Plugging this back in the previous equality we get the result, at least when f is in $\mathcal{C}^2$. □

Theorem 1.18 (logarithmic Sobolev inequality)**.** *Let f be a non negative, $\mathcal{C}^1$–smooth, integrable function. The following inequality holds true:*

$$\mathrm{Ent}_{\gamma_n}(f) \leq \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \,\mathrm{d}\gamma_n. \tag{1.9}$$

Exercise (Optimality)**.** *Check on functions f of the form $f(x) = \mathrm{e}^{\langle u,x\rangle}$ for some vector u in $\mathbb{R}^n$ that the inequality is sharp.*

Proof. Assume that f is in $\mathcal{C}^2$ and satisfies $f \geq \epsilon$ for some positive ϵ. Then we can apply (1.6) with the function $\phi = x \log x$. We obtain

$$\mathrm{Ent}_{\gamma_n}(f) = \int_0^{+\infty} \int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f} \, \mathrm{d}\gamma_n \mathrm{d}t$$

Then using (1.8) and stationarity, we get the result. □

Let us conclude this section by showing that the Poincaré inequality can formally be derived from the log–Sobolev inequality. Let h be a bounded, $\mathcal{C}^1$–smooth function. If ϵ is small enough then $1 + \epsilon h$ is non negative. Since

$$(1+t)\log(1+t) = t + \frac{t^2}{2} + o(t^2)$$

as t tends to 0 we easily get

$$\mathrm{Ent}_{\gamma_n}(1+\epsilon h) = \frac{\epsilon^2}{2} \mathrm{Var}_{\gamma_n}(h) + o(\epsilon^2),$$

where $\mathrm{Var}_{\gamma_n}(h)$ denotes the variance of h under the Gaussian measure:

$$\mathrm{Var}_{\gamma_n}(h) = \int_{\mathbb{R}^n} h^2 \, \mathrm{d}\gamma_n - \left(\int_{\mathbb{R}^n} h \, \mathrm{d}\gamma_n \right)^2 .$$

Similarly

$$\int_{\mathbb{R}^n} \frac{|\nabla(1+\epsilon h)|^2}{1+\epsilon h} \, \mathrm{d}\gamma_n = \epsilon^2 \int_{\mathbb{R}^n} |\nabla h|^2 \, \mathrm{d}\gamma_n + o(\epsilon^2).$$

Therefore, applying (1.9) to the function $f = 1 + \epsilon h$ and sending ϵ to 0, we obtain the Poincaré inequality for h.

Remark (Alternative proof)**.** *Let us give an alternative proof of the Poincaré and logarithmic Sobolev inequalities. Fix $t \geq 0$, $x \in \mathbb{R}^d$, $f : \mathbb{R}^d \to I$, and define*

$$s \in [0,t] \mapsto \beta(s) = P_s(\phi(P_{t-s} f)) \quad \textit{where} \quad \phi(x) = x^2 \quad \textit{or} \quad \phi(x) = x \log x.$$

Note that we dropped the x in the notation, namely $P_s(\cdot) = P_s(\cdot)(x)$. We have

$$\beta(t) - \beta(0) = P_t(\phi(f)) - \phi(P_t f) = \mathbb{E}^{\phi}_{P_t(\cdot)}(f).$$

Here $P_t(\cdot) = P_t(\cdot)(x)$ is a probability measure defined by $\mathbb{E}_{P_t(\cdot)(x)}(f) = P_t(f)(x)$. On the other hand, setting $g = P_{t-s} f$, a computation reveals that

$$\beta'(s) = P_s(L(\phi(g)) - Lg\phi'(g)).$$

Now a direct computation gives $L(\phi(g)) - Lg\phi'(g) = \phi''(g)|\nabla g|^2$, and thus

$$\beta'(s) = P_s(\phi''(P_{t-s} f)|\nabla P_{t-s} f|^2).$$

Mehler's formula gives the sub-commutation $|\nabla P_{t-s} f| \leq \mathrm{e}^{-(t-s)} P_{t-s}(|\nabla f|)$ and therefore

$$\beta'(s) \leq \mathrm{e}^{-2(t-s)} P_s(\phi''(P_{t-s} f) P_{t-s}(|\nabla f|)^2).$$

Jensen's inequality for the convex function $(u,v) \mapsto \phi''(u)v^2$ gives

$$\phi''(P_{t-s} f) P_{t-s}(|\nabla f|)^2 \leq P_{t-s}(\phi''(f)|\nabla f|^2).$$

Therefore, we obtain

$$\beta'(s) \leq \mathrm{e}^{-2(t-s)} P_t(\phi''(f)|\nabla f|^2).$$

By integrating on $[0,t]$ *we get that for any* $t \geq 0$ *and* $x \in \mathbb{R}$*, the probability measure* $P_t(\cdot)(x) = \mathcal{N}(x\mathrm{e}^{-t}, 1-\mathrm{e}^{-2t})$ *satisfies, for any* $\mathcal{C}^2$ *test function* f*,*

$$P_t(\phi(f))(x) - \phi(P_t(f)(x)) \leq \frac{1-\mathrm{e}^{-2t}}{2} P_t(\phi''(f)|\nabla f|^2)(x).$$

in particular a Poincaré and a logarithmic Sobolev inequality with constants $(1-\mathrm{e}^{-2t})$ *and* $(1-\mathrm{e}^{-2t})/2$ *respectively. By sending t to infinity or by using a translation and a dilation we can get these inequalities for* $\gamma_n = \mathcal{N}(0, I_n)$ *with (optimal) constants* 2 *and* 1 *respectively.*

1.4 Convergence to equilibrium

We have seen that $X_t \to \gamma_n$ in law as t tends to $+\infty$. We shall see now that the Poincaré inequality and the logarithmic Sobolev inequalities allow to quantify this convergence.

Theorem 1.19. *For every* $f \in \mathbb{L}^2(\gamma_n)$ *we have*

$$\mathrm{Var}_{\gamma_n}(P_t f) \leq \mathrm{e}^{-2t} \mathrm{Var}_{\gamma_n}(f).$$

For every non negative integrable f *we have*

$$\mathrm{Ent}_{\gamma_n}(P_t f) \leq \mathrm{e}^{-2t} \mathrm{Ent}_{\gamma_n}(f).$$

Proof. Let $\alpha(t) = \mathrm{Var}(P_t f)$. By stationarity

$$\alpha(t) = \int_{\mathbb{R}^n} (P_t f)^2 \,\mathrm{d}\gamma_n - \left(\int_{\mathbb{R}^n} f \mathrm{d}\gamma_n\right)^2.$$

As we have seen before

$$\alpha'(t) = \frac{d}{\mathrm{d}t}\left(\int_{\mathbb{R}^n} (P_t f)^2 \,\mathrm{d}\gamma_n\right) = -2\int_{\mathbb{R}^n} |\nabla P_t f|^2 \,\mathrm{d}\gamma_n$$

So applying Poincaré inequality to $P_t f$ we obtain $\alpha'(t) \leq -2\alpha(t)$ for every $t \geq 0$. By Gronwall's lemma we get $\alpha(t) \leq \mathrm{e}^{-2t}\alpha(0)$ which is the first inequality. For the second inequality, let

$$\beta(t) = \mathrm{Ent}_{\gamma_n}(P_t f)$$

and observe that the logarithmic Sobolev inequality yields $\beta'(t) \leq -2\beta$ which gives the result by Gronwall's lemma. Note that the decay takes the form $\mathrm{e}^{-\frac{2}{c_{\mathrm{PI}}}t}$ for the variance and $\mathrm{e}^{-\frac{4}{c_{\mathrm{LSI}}}t}$ for the entropy, with $c_{\mathrm{PI}} = 1$ and $c_{\mathrm{LSI}} = 2$ here for γ_n. □

Lemma 1.20 (Partial differential equation)**.** *For any* $t \geq 0$*, let* μ_t *be the law of* X_t*. If* μ_0 *has density* h_0 *with respect to* γ_n *then* μ_t *has density* $h_t = P_t^*(h_0)$ *with respect to* γ_n*, where* P_t^* *is the transpose of* P_t *in* $\mathbb{L}^2(\gamma_n)$*. Since* γ_n *is reversible,* $P_t^* = P_t$ *and*

$$\partial_t h_t(x) = \partial_t P_t(h_0) = L P_t(h_0) = L h_t(x) = \Delta_x h_t(x) - \langle \nabla_x h_t(x), x\rangle.$$

How about densities with respect to the Lebesgue measure? If φ_n *is the density of* γ_n *with respect to the Lebesgue measure, then* μ_t *has density* $g_t = h_t\varphi_n$ *with respect to the Lebesgue measure, and is solution of the Fokker-Planck partial differential equation given by*

$$\partial_t g_t(x) = \Delta_x g_t(x) + \mathrm{div}_x(x g_t(x)).$$

This corresponds to the transpose of the semigroup in $\mathbb{L}^2(dx)$ *instead of* $\mathbb{L}^2(\gamma_n)$*. Moreover, the equilibrium corresponds in a way to* $h_\infty = 1$ *and to* $g_\infty = \varphi_n$ *respectively.*

Proof. Exercise! □

Reformulate in terms of convergence of measure:

$$\chi^2(\mu P_t \mid \gamma_n) \leq \mathrm{e}^{-2t}\chi^2(\mu \mid \gamma_n)$$
$$\mathrm{H}(\mu P_t \mid \gamma_n) \leq \mathrm{e}^{-2t}\mathrm{H}(\mu \mid \gamma_n).$$

Here $\chi^2(\nu \mid \mu)$ denotes the chi-square divergence defined by

$$\chi^2(\nu \mid \mu) := \mathrm{Var}_\mu(f) \quad \text{where} \quad f := \frac{\mathrm{d}\nu}{\mathrm{d}\mu};$$

while H denotes the relative entropy or Kullback-Leibler divergence defined for any probability measures μ and ν by

$$\mathrm{H}(\nu \mid \mu) := \mathrm{Ent}_\mu(f) = \int f \log f \,\mathrm{d}\mu \quad \text{where} \quad f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

It is customary to set $\chi^2(\nu \mid \mu) = +\infty$ and $\mathrm{H}(\nu \mid \mu) = +\infty$ if ν is not absolutely continuous with respect to μ, or when $f \notin \mathbb{L}^2(\mu)$ or $f \log f \notin \mathbb{L}^1(\mu)$ respectively.

1.5 Amnesia and long time behavior

We can measure the way that the Ornstein–Uhlenbeck process forgets its initial position along the time by using for instance the relative entropy between the law of the process at time t started from two different positions. It is also customary to use the Wasserstein distance instead of the relative entropy. Recall that the Wasserstein distance of order 2 between two probability measures μ_1 and μ_2 is

$$\mathrm{W}_2(\mu_1, \mu_2) = \inf_{\substack{X_1 \sim \mu_1 \\ X_2 \sim \mu_2}} \sqrt{\mathbb{E}(|X_1 - X_2|^2)},$$

where the infimum runs over the couples (X_1, X_2) with marginals μ_1 and μ_2.

Theorem 1.21 (Relative entropy and Wasserstein distance for Gaussians)**.** *If $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$ and $\mu_2 = \mathcal{N}(m_2, \Sigma_2)$ on $\mathbb{R}^n$ then*

$$\mathrm{W}_2(\mu_1, \mu_2)^2 = |m_1 - m_2|^2 + \mathrm{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}),$$

and in particular $\mathrm{W}_2(\mu_1, \mu_2)^2 = |m_1 - m_2|^2 + \mathrm{Tr}((\sqrt{\Sigma_1} - \sqrt{\Sigma_2})^2)$ when Σ_1 and Σ_2 commute. Moreover if Σ_1 et Σ_2 are invertible then

$$\mathrm{H}(\mu_1 \mid \mu_2) = \frac{1}{2}\Big(\log \frac{\det \Sigma_2}{\det \Sigma_1} + \mathrm{Tr}(\Sigma_2^{-1}\Sigma_1) - n + (m_1 - m_2)^T \Sigma_2^{-1}(m_1 - m_2)\Big).$$

In dimension $n = 1$, if $\mu_1 = \mathcal{N}(m_1, \sigma_1)$ and $\mu_2 = \mathcal{N}(m_2, \sigma_2)$ on $\mathbb{R}$ then

$$\mathrm{W}_2(\mu_1, \mu_2)^2 = (m_1 - m_2)^2 + (\sigma_1 - \sigma_2)^2,$$

while if $\sigma_1 > 0$ and $\sigma_2 > 0$ then

$$\mathrm{H}(\mu_1 \mid \mu_2) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 - \sigma_2^2 + (m_1 - m_2)^2}{2\sigma_2^2}.$$

Proof when $n = 1$. Let us start with the formula for W_2. If (X_1, X_2) is a couple of random variables such that $X_i \sim \mu_i = \mathcal{N}(m_i, \sigma_i^2)$ for $i = 1, 2$. We have

$$\begin{aligned}\mathbb{E}(|X_1 - X_2|^2) &= \mathbb{E}(X_1^2) + \mathbb{E}(X_2)^2 - 2\mathbb{E}(X_1 X_2)\\ &= (m_1 - m_2)^2 + \sigma_1^2 + \sigma_2^2 - 2\mathrm{Cov}(X_1, X_2).\end{aligned}$$

This formula depends only on the mean and on the covariance matrix of the random vector (X_1, X_2). We have

$$\mathrm{W}_2(\mu_1, \mu_2)^2 = (m_1 - m_2)^2 + \sigma_1^2 + \sigma_2^2 - 2 \sup_{C \in \mathcal{C}} C_{12}$$

where $\mathcal{C}$ is the set of 2×2 covariance matrices with a diagonal prescribed by $C_{11} = \sigma_1^2$ and $C_{22} = \sigma_2^2$. Since the set of covariance matrices coincidees with the set of symmetric matrices with non-negative spectrum, $C \in \mathcal{C}$ gives the constraint $\det(C) = \sigma_1^2\sigma_2^2 - C_{12}^2 \geq 0$, hence $C_{12} = \sigma_1\sigma_2$, and we are done.

For the entropy formula, we write,

$$\begin{aligned}\mathrm{H}(\mu_1 \mid \mu_2) &= \int \log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}\,\mathrm{d}\mu_1\\ &= \int \left(\log \frac{\sigma_2}{\sigma_1} - \frac{(x - m_1)^2}{2\sigma_1^2} + \frac{(x - m_2)^2}{2\sigma_2^2}\right) \mu_1(dx)\\ &= \log \frac{\sigma_2}{\sigma_1} - \frac{1}{2} + \frac{\sigma_1^2 + m_1^2 - 2m_1m_2 + m_2^2}{2\sigma_2^2}\\ &= \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 - \sigma_2^2 + (m_1 - m_2)^2}{2\sigma_2^2}.\end{aligned}$$

□

Since the Ornstein–Uhlenbeck process is a Gaussian process, we may use the preceding formulas to quantify the long time behavior when the initial condition is itself Gaussian. Namely for any $x_1, x_2 \in \mathbb{R}$ and any $t \geq 0$, if $\mu_1 = P_t(\cdot)(x_1) = \mathcal{N}(x_1\mathrm{e}^{-t}, 1 - \mathrm{e}^{-2t})$ and $\mu_2 = P_t(\cdot)(x_2) = \mathcal{N}(x_2\mathrm{e}^{-t}, 1 - \mathrm{e}^{-2t})$,

$$\mathrm{W}_2(P_t(\cdot)(x_1), P_t(\cdot)(x_2))^2 = \mathrm{e}^{-2t}(x_1 - x_2)^2 \underset{t\to\infty}{\searrow} 0$$

and

$$\mathrm{H}(P_t(\cdot)(x_1) \mid P_t(\cdot)(x_2)) = \frac{\mathrm{e}^{-2t}(x_1 - x_2)^2}{2(1 - \mathrm{e}^{-2t})} \underset{t\to\infty}{\searrow} 0.$$

This can be extended to more general initial distributions, and the case of the Wasserstein distance is particularly simple due to its relation with coupling.

Theorem 1.22 (Convergence in Wasserstein distance)**.** *If μ_0 and μ_0' are probability measures with finite second moment, and if μ_t (respectively μ_t') is the law of X_t (respectively X_t') when $X_0 \sim \mu_0$ (respectively $X_0' \sim \mu_0'$), then for any $t \geq 0$,*

$$\mathrm{W}_2(\mu_t, \mu_t') \leq \mathrm{e}^{-t}\mathrm{W}_2(\mu_0, \mu_0').$$

In particular if $\mu_0' = \gamma_1$ then for any $t \geq 0$,

$$\mathrm{W}_2(\mu_t, \gamma_1) \leq \mathrm{e}^{-t}\mathrm{W}_2(\mu_0, \gamma).$$

Proof. Let (X_0, X_0') be a coupling of μ_0 and μ_0' independent of a Brownian motion $B = (B_t)_{t\geq 0}$. We construct two processes $(X_t)_{t\geq 0}$ and $(X_t')_{t\geq 0}$ with respective initial conditions X_0 and X_0' and driven by the same Brownian motion B. We say that the processes are coupled. We have then

$$X_t - X_t' = X_0 - X_0' - \int_0^t (X_s - X_s')\mathrm{d}s.$$

It follows that

$$|X_t - X_t'|^2 = \mathrm{e}^{-2t}|X_0 - X_0'|^2.$$

By definition of $\mathrm{W}_2(\mu_t, \mu_t')$, we obtain

$$\mathrm{W}_2(\mu_t, \mu_t')^2 \leq \mathrm{e}^{-2t}\mathbb{E}(|X_0 - X_0'|^2).$$

It remains to take the infimum over all couplings of μ_0 and μ_0'. □

1.6 Tensorization and Central Limit Theorem

This section is devoted to the tensorization property of the variance and the entropy. This property is at the heart of the dimension free nature of the Poincaré and the logarithmic Sobolev inequalities. It allows to provide a proof of the Poincaré and of the logarithmic Sobolev inequalities for the Gaussian measure by using the Central Limit Theorem, starting from elementary inequalities on the two-point space.

The Gaussian measure appears as a limiting distribution in the asymptotic analysis of product spaces, due to the central limit phenomenon. The simplest product space is the discrete cube $\{0,1\}^n$ equipped with the product Bernoulli probability measure $\mu_n = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes n}$, which is the uniform probability measure. This model is called "the two-point-space" when $n = 1$. Thanks to the Central Limit Theorem states, if $X_n = (X_{n,1}, \ldots, X_{n,n}) \sim \mu_n$ for every n then the $X_{n,i}$ are i.i.d. with mean $m = 1/2$ and variance $\sigma^2 = 1/4$, and therefore

$$\frac{X_{n,1} + \cdots + X_{n,n} - nm}{\sqrt{n\sigma^2}} = \frac{2(X_{n,1} + \cdots + X_{n,n}) - n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} \gamma_1.$$

In other words, for any continuous and bounded $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\{0,1\}^n} g_n \,\mathrm{d}\mu_n \underset{n\to\infty}{\longrightarrow} \int f \,\mathrm{d}\gamma_1 \quad \text{where} \quad g_n(x) := f\left(\frac{2(x_1 + \cdots + x_n) - n}{\sqrt{n}}\right).$$

Theorem 1.23 (Tensorization). *Let $(E_1, \mathcal{A}_1, \mu_1), \ldots, (E_n, \mathcal{A}_n, \mu_n)$ be probability spaces. Let $\mu_1 \otimes \cdots \otimes \mu_n$ be the product probability measure on $(E_1 \times \cdots \times E_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n)$. Let $\phi : \mathcal{I} \to \mathbb{R}$ be convex and such that $(u, v) \mapsto \phi''(u)v^2$ is convex. Then, for any $f : E_1 \times \cdots \times E_n \to \mathbb{R}$ such that $\phi(f) \in \mathbb{L}^1(\mu_1 \otimes \cdots \otimes \mu_n)$,*

$$\mathbb{E}_\mu^\phi(f) \leq \sum_{i=1}^n \mathbb{E}_\mu \mathbb{E}_{\mu_i}^\phi(f),$$

where the subscript μ_i indicates that the integration concerns the i-th variable only.

For $\phi(u) = u^2$ on $\mathcal{I} = \mathbb{R}$ we get the variance and the result reads

$$\mathrm{Var}_{\mu_1\otimes\cdots\otimes\mu_n}(f) \leq \mathbb{E}_{\mu_1\otimes\cdots\otimes\mu_n}(\mathrm{Var}_{\mu_1}(f) + \cdots + \mathrm{Var}_{\mu_n}(f)) \tag{1.10}$$

while for $\phi(u) = u\log(u)$ on $\mathcal{I} = [0, \infty)$ we get the entropy and the result reads

$$\mathrm{Ent}_{\mu_1\otimes\cdots\otimes\mu_n}(f) \leq \mathbb{E}_{\mu_1\otimes\cdots\otimes\mu_n}(\mathrm{Ent}_{\mu_1}(f) + \cdots + \mathrm{Ent}_{\mu_n}(f)). \tag{1.11}$$

Proof. By induction on n we only have to consider the case $n = 2$, for which the desired bound boils down after expansion and rearrangement of terms to

$$\mathbb{E}^{\phi}_{\mu_2}(\mathbb{E}_{\mu_1}(f)) \leq \mathbb{E}_{\mu_1}(\mathbb{E}^{\phi}_{\mu_2}(f)).$$

In the case of the variance this follows from the Cauchy–Schwartz inequality[2]. The general proof is based on convexity. The ϕ-entropy functional

$$f \mapsto \mathbb{E}^{\phi}_{\mu}(f) := \mathbb{E}_{\mu}(\phi(f)) - \phi(\mathbb{E}_{\mu} f)$$

is convex as soon as $C^{\phi} : (u, v) \mapsto \phi''(u)v^2$ is convex, which is the case for the variance case $\phi(u) = u^2$ and the entropy case $\phi(u) = u \log(u)$. Indeed, since convexity is a unidimensional property, if $\alpha(t) := \mathbb{E}^{\phi}_{\mu}(f + tg)$ then $\alpha''(0) = \mathbb{E}_{\mu}(C^{\phi}(f, g)) - C^{\Phi}(\mathbb{E}_{\mu}(f), \mathbb{E}_{\mu}(g)) \geq 0$ by the Jensen inequality. As a consequence, the functional $\mathbb{E}^{\phi}_{\mu}$ is equal to the enveloppe of its directional tangents[3], namely the following variational formula holds[4], with equality when $f = g$,

$$\mathbb{E}^{\phi}_{\mu}(f) = \sup_{g : \phi(g) \in \mathbb{L}^1(\mu)} \{\mathbb{E}^{\phi}_{\mu}(g) + \mathbb{E}_{\mu}((\phi'(g) - \phi'(\mathbb{E}_{\mu} g))(f - g)\}.$$

By using this variational formula for μ_2, we obtain

$$\begin{aligned}
\mathbb{E}^{\phi}_{\mu_2}(\mathbb{E}_{\mu_1}(f)) &= \sup_g \{\mathbb{E}^{\phi}_{\mu_2}(g) + \mathbb{E}_{\mu_2}(\phi'(g) - \phi'(\mathbb{E}_{\mu_2} g))(\mathbb{E}_{\mu_1} f - g)\} \\
&= \sup_g \mathbb{E}_{\mu_1} \{\mathbb{E}^{\phi}_{\mu_2}(g) + \mathbb{E}_{\mu_2}(\phi'(g) - \phi'(\mathbb{E}_{\mu_2} g))(f - g)\} \\
&\leq \mathbb{E}_{\mu_1} \{\sup_g \{\mathbb{E}^{\phi}_{\mu_2}(g) + \mathbb{E}_{\mu_2}(\phi'(g) - \phi'(\mathbb{E}_{\mu_2} g))(f - g)\}\} \\
&= \mathbb{E}_{\mu_1}(\mathbb{E}^{\phi}_{\mu_2}(f))
\end{aligned}$$

where the suprema are taken over functions $g : E_2 \to \mathcal{I}$ such that $\phi(g) \in \mathbb{L}^1(\mu_2)$. □

Let us prove now the Poincaré inequality for γ_1. Let $\mu_n = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes n}$ be the uniform distribution on the cube $\{0, 1\}^n$. For any $g : \{0, 1\} \to \mathbb{R}$,

$$\mathrm{Var}_{\mu_1}(g) = \frac{g(0)^2 + g(1)^2}{2} - \Big(\frac{g(0) + g(1)}{2}\Big)^2 = \frac{(g(1) - g(0))^2}{4}.$$

Using the tensorization property, for any $n \geq 1$ and $g : \{0, 1\}^n \to \mathbb{R}$,

$$\mathrm{Var}_{\mu_n}(g) \leq \frac{1}{4}\mathbb{E}_{\mu_n}((\mathrm{D}_1 g)^2 + \cdots + (\mathrm{D}_n g)^2)$$

where $(\mathrm{D}_i g)^2(x) := (g(x + e_i) - g(x))^2$ where $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{R}^n$. Now, let $f : \mathbb{R} \to \mathbb{R}$ be $\mathcal{C}^2$ and compactly supported, and set

$$g(x) = g_n(x) = f(s_n(x)) \quad \text{with} \quad s_n(x) := \frac{2(x_1 + \cdots + x_n) - n}{\sqrt{n}}.$$

Now, by using a Taylor formula, for any $i = 1, \ldots, n$ and $x \in \{0, 1\}^n$,

$$(\mathrm{D}_i g)^2(x) = \Big(\frac{2}{\sqrt{n}} f'(s_n(x)) + o\Big(\frac{1}{\sqrt{n}}\Big)\Big)^2 = \frac{4}{n} f'^2(s_n(x)) + o\Big(\frac{1}{n}\Big)$$

[2] $\mathrm{Var}_{\mu_2}(\mathbb{E}_{\mu_1} f) = \mathbb{E}_{\mu_2}((\mathbb{E}_{\mu_1} f - \mathbb{E}_{\mu_2}\mathbb{E}_{\mu_1} f)^2) \leq \mathbb{E}_{\mu_2}\mathbb{E}_{\mu_1}((f - \mathbb{E}_{\mu_2} f)^2) = \mathbb{E}_{\mu_1}\mathrm{Var}_{\mu_2}(f)$.

[3] Namely $c(0) = \sup_{s \in [0,1]}\{c(s) + c'(s)(0 - s)\}$ where $c(s) = \mathbb{E}^{\phi}_{\mu}(f + s(g - f))$ valid for any g.

[4] Variance case rewrites $\mathrm{Var}_{\mu}(f) = \sup\{\mathrm{Cov}_{\mu}(f, g) - 2\mathrm{Var}_{\mu}(g) : g\}$, supremum achieved for $g = f$, while the entropy case rewrites $\mathrm{Ent}_{\mu}(f) = \sup\{\mathbb{E}_{\mu}(fh) : \mathbb{E}_{\mu}(h) \leq 1\}$ and the supremum is achieved for $h = \log(f)$.

where the o is uniform in x since f is $\mathcal{C}^2$ and compactly supported. Now, the Central Limit Theorem yields

$$\mathrm{Var}_{\gamma_1}(f) = \lim_{n\to\infty} \mathrm{Var}_{\mu_n}(g) \leq \lim_{n\to\infty} \mathbb{E}_{\mu_n}(f'^2(s_n)) = \mathbb{E}_{\gamma_1}(f'^2).$$

This is the optimal Poincaré inequality for γ_1. To pass from γ_1 to $\gamma_n = \gamma_1^{\otimes n}$, we can use the tensorization property (1.10) again to get, for a sufficiently regular $f : \mathbb{R}^n \to \mathbb{R}$, thanks to the fact that $|\nabla f|^2 := \sum_{i=1}^n (\mathrm{D}_i f)^2$,

$$\mathrm{Var}_{\gamma_n}(f) \leq \mathbb{E}_{\gamma_n}(|\nabla f|^2).$$

Let us explain now how to enlarge the class of test functions. This can be done by usual approximation procedures. If for instance $f \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ is such that $f \in \mathbb{L}^2(\gamma_n)$ and $\nabla f \in \mathbb{L}^2(\gamma_n)$, and if $\eta_n \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ with $\mathbf{1}_{x\in\mathbb{R}^n:|x|\leq n} \leq \eta_n \leq \mathbf{1}_{x\in\mathbb{R}^n:|x|\leq n+1}$, then, using the Fatou lemma, the Poincaré inequality for $f\eta_n$, and dominated convergence, we get

$$\int f^2 \mathrm{d}\gamma_n \leq \varliminf_{n\to\infty} \int (f\eta_n)^2 \mathrm{d}\gamma_n \leq \varliminf_{n\to\infty} \int |\eta_n \nabla f + f \nabla \eta_n|^2 \mathrm{d}\gamma_n = \int |\nabla f|^2 \mathrm{d}\gamma_n.$$

Further approximation arguments provide Poincaré inequality for $f \in \mathbb{H}^1(\gamma_1) = \mathbb{W}^{1,2}(\gamma_n)$.

Remark (Integrability). *It is easy to check that if $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ is such that $f' \in \mathbb{L}^2(\gamma_1)$ then $f \in \mathbb{L}^2(\gamma_1)$. Indeed, assuming without loss of generality that $f(x) = 0$ if $x < 0$ and $f(0) = 0$, we have, using the Cauchy–Schwarz inequality and the Fubini–Tonelli theorem,*

$$\begin{aligned}
\int_0^\infty |f(x)|^2 \mathrm{d}\gamma_1(x) &= \int_0^\infty \Big| \int_0^x f'(y)\mathrm{d}y \Big|^2 \mathrm{d}\gamma_1(x) \\
&\leq \int_0^\infty x \Big(\int_0^x |f'(y)|^2 \mathrm{d}y \Big) \gamma_1(\mathrm{d}x) \\
&= \int_0^\infty |f'(y)|^2 \Big(\int_y^\infty x\gamma_1(\mathrm{d}x) \Big) \mathrm{d}y \\
&= \int_0^\infty |f'(y)|^2 \gamma_1(\mathrm{d}y).
\end{aligned}$$

For the logarithmic Sobolev inequality, we can proceed exactly as we did above for the Poincaré inequality. The starting point is the following inequality on the two-points space: for any $g : \{0, 1\} \to \mathbb{R}$,

$$\mathrm{Ent}_{\mu_1}(g^2) \leq \frac{(g(1) - g(0))^2}{2},$$

which reads, with $a := g(0)$ and $b := g(1)$, as the sharp inequality

$$\frac{a^2 \log(a^2) + b^2 \log(b^2)}{2} - \frac{a^2 + b^2}{2} \log \frac{a^2 + b^2}{2} \leq \frac{(a - b)^2}{2}.$$

By homogeneity, this reduces to the even simpler elementary inequality

$$u \log(u) + (2 - u) \log(2 - u) \leq (\sqrt{u} - \sqrt{2 - u})^2, \quad 0 \leq u \leq 2.$$

This strategy of proof of the logarithmic Sobolev inequality for γ_n using the Central Limit Theorem goes back to the seminal work [19] of Leonard Gross.

Chapter 2

Hypercontractivity, spectral gap, information theory

This chapter is devoted first to a couple of important reformulations: hypercontractivity as a reformulation of the logarithmic Sobolev inequality, and spectral gap via Hermite polynomials as a reformulation of the Poincaré inequality. A link is made with the formula of the Gaussian Unitary Ensemble via the quantum harmonic oscillator. The third and last part of the chapter is devoted to a study of an Euclidean form of the logarithmic Sobolev inequality, in information theory, via Shannon entropy and Fisher information.

2.1 Hypercontractivity

In this section we shall show that the semigroup (P_t) is hypercontractive: if f belongs to $\mathbb{L}^p(\gamma_n)$ for some $p > 1$ then $P_t f$ belongs to $\mathbb{L}^q(\gamma_n)$ for some $q > p$. The precise result states as follows.

Theorem 2.1 (Hypercontractivity (Nelson))**.** *Let $p > 1$ and let $t > 0$ and set $p(t) = 1 + (p-1)\mathrm{e}^{2t}$. Observe that $p(t) > p$. Then for every $f \in \mathbb{L}^p(\gamma_n)$*

$$\|P_t f\|_{p(t)} \leq \|f\|_p.$$

In other words the operator P_t is bounded from $\mathbb{L}^p(\gamma_n)$ to $\mathbb{L}^{p(t)}(\gamma_n)$ and has norm 1. Moreover, if $q > p(t)$ then P_t is not even bounded from $\mathbb{L}^p(\gamma_n)$ to $\mathbb{L}^q(\gamma_n)$.

The proof makes use of the logarithmic Sobolev inequality. The basic idea is that the derivative of the $\mathbb{L}^p$ norm with respect to p will bring the entropy, while the derivative of the semigroup P_t with respect to t will bring the generator.

Note that the critical exponent $p(t)$ in hypercontractivity above does not depend on the dimension n, just like the logarithmic Sobolev constant $1/2$ in front of its right hand side.

Proof. One can assume that $f \geq 0$ since $|P_t f| \leq P_t|f|$. Set $\alpha(t) = \log\|P_t f\|_{p(t)}$. To lighten the notation, let us set $f_t = P_t f$. We have, for any $t \geq 0$,

$$\begin{aligned}
\alpha'(t) &= \left(\frac{1}{p(t)}\log\int f_t^{p(t)}\,\mathrm{d}\gamma_n\right)' \\
&= -\frac{p'(t)}{p(t)^2}\log\int f_t^{p(t)}\,\mathrm{d}\gamma_n + \frac{1}{p(t)}\frac{\left(\int f_t^{p(t)}\,\mathrm{d}\gamma_n\right)'}{\int f_t^{p(t)}\,\mathrm{d}\gamma_n}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{p'(t)}{p(t)^2}\log\int f_t^{p(t)}\,\mathrm{d}\gamma_n + \frac{1}{p(t)}\frac{\displaystyle\int\left(p'(t)\log f_t + p(t)\frac{Lf_t}{f_t}\right)f_t^{p(t)}\,\mathrm{d}\gamma_n}{\displaystyle\int f_t^{p(t)}\,\mathrm{d}\gamma_n}\\
&= -\frac{p'(t)}{p(t)^2}\log\int f_t^{p(t)}\,\mathrm{d}\gamma_n + \frac{p'(t)}{p(t)^2}\frac{\displaystyle\int f_t^{p(t)}\log f_t^{p(t)}\,\mathrm{d}\gamma_n}{\displaystyle\int f_t^{p(t)}\,\mathrm{d}\gamma_n} + \frac{\displaystyle\int (Lf_t)f_t^{p(t)-1}\,\mathrm{d}\gamma_n}{\displaystyle\int f_t^{p(t)}\,\mathrm{d}\gamma_n}\\
&= \frac{p'(t)}{p(t)^2}\frac{1}{\displaystyle\int f_t^{p(t)}\,\mathrm{d}\gamma_n}\left(\mathrm{Ent}_{\gamma_n}(f_t^{p(t)}) + \frac{p(t)^2}{p'(t)}\int (Lf_t)f_t^{p(t)-1}\,\mathrm{d}\gamma_n\right).
\end{aligned}
$$

Now the logarithmic Sobolev inequality and the integration by parts give

$$
\begin{aligned}
\mathrm{Ent}_{\gamma_n}(g^p) &\leq \frac{1}{2}\int\frac{|\nabla g^p|^2}{g^p}\,\mathrm{d}\gamma_n\\
&= \frac{p^2}{2}\int|\nabla g|^2 g^{p-2}\,\mathrm{d}\gamma_n\\
&= \frac{p^2}{2(p-1)}\int\left\langle\nabla g, \nabla g^{p-1}\right\rangle\,\mathrm{d}\gamma_n\\
&= -\frac{p^2}{2(p-1)}\int(Lg)g^{p-1}\,\mathrm{d}\gamma_n.
\end{aligned}
$$

Using this inequality for $g = f_t$ and $p = p(t)$, and using $2(p(t)-1) = p'(t)$, we obtain that $\alpha'(t) \leq 0$ for any $t \geq 0$, and as a consequence

$$
\log\|P_t f\|_{p(t)} = \alpha(t) \leq \alpha(0) = \log\|f\|_p.
$$

Finally, if now $q > p(t)$ then taking $f_\lambda(x) = \mathrm{e}^{\langle\lambda,x\rangle}$ for some parameter $\lambda \in \mathbb{R}^n$ gives

$$
\|f_\lambda\|_p = \mathrm{e}^{\frac{1}{2}p|\lambda|^2} \quad\text{and}\quad P_t f_\lambda = \mathrm{e}^{\frac{1}{2}|\lambda|^2(1-\mathrm{e}^{-2t})}f_{\lambda\mathrm{e}^{-t}}
$$

and therefore

$$
\frac{\|P_t f_\lambda\|_q}{\|f_\lambda\|_p} = \mathrm{e}^{\frac{1}{2}\lambda^2(\mathrm{e}^{-2t}(q-1)+1-p)},
$$

a quantity which tends to $+\infty$ as $\lambda \to \pm\infty$ since $q > p(t) = 1 + (p-1)\mathrm{e}^{2t}$. □

Remark (Hypercontractivity and logarithmic Sobolev inequality)**.** *This proof show more generally that a semigroup satisfying the logarithmic Sobolev inequality is hypercontractive. Moreover, it is pretty clear from the argument that the implication can be reversed and that log–Sobolev and hypercontractivity are equivalent. This equivalence between hypercontractivity and logarithmic Sobolev inequality is due to Leonard Gross [19].*

Let us rephrase the latter result in a slightly different way. Let $p, q \geq 1$, let $t \geq 0$. Observe that by duality

$$
\|P_t f\|_q = \sup_g\left\{\frac{\displaystyle\int_{\mathbb{R}^n}(P_t f)g\,\mathrm{d}\gamma_n}{\|g\|_{q'}}\right\}
$$

where $q' = 1/(1-1/q) = q/(q-1)$ is the conjugate exponent of q. Observe also that if (X_t) is an Ornstein–Uhlenbeck process in equilibrium then

$$
\begin{aligned}
\int_{\mathbb{R}^n}(P_t f)g\,\mathrm{d}\gamma_n &= \mathbb{E}[P_t f(X_0)g(X_0)]\\
&= \mathbb{E}\left[\mathbb{E}[f(X_t) \mid X_0]g(X_0)\right] = \mathbb{E}[f(X_t)g(X_0)].
\end{aligned}
$$

Hypercontractivity thus asserts that if $q \le 1 + (p-1)\mathrm{e}^{2t}$ then

$$\mathbb{E}[f(X_t)g(X_0)] \le \mathbb{E}[|f(X_t)|^p]^{1/p}\mathbb{E}[|g(X_0)|^{q'}]^{1/q'},$$

for every f, g. Equivalently

$$\mathbb{E}[F(X_t)^\alpha G(X_0)^\beta] \le \mathbb{E}[F(X_t)]^\alpha \mathbb{E}[G(X_0)]^\beta,$$

for every $F, G \ge 0$, where $\alpha = 1/p$ and $\beta = 1/q'$. Recall that (X_0, X_t) is a centered Gaussian vector with covariance

$$\begin{pmatrix} I_n & \rho\, I_n \\ \rho\, I_n & I_n \end{pmatrix}$$

where $\rho = \mathrm{e}^{-t}$ and and note that the hypothesis $q \le 1 + (p-1)\mathrm{e}^{2t}$ reads $\rho^2 \le (1-1/\alpha)(1-1/\beta)$ in terms of ρ, α, β. Therefore the hypercontractivity property of the Ornstein–Uhlenbeck semigroup can be reformulated as follows.

Theorem 2.2. *Let $\rho, \alpha, \beta \in [0,1]$ and let (X, Y) be a Gaussian vector on $\mathbb{R}^n$ centered at 0 and having covariance matrix*

$$\begin{pmatrix} I_n & \rho\, I_n \\ \rho I_n & I_n \end{pmatrix}.$$

If $\rho^2 \le (1-1/\alpha)(1-1/\beta)$ then for every non–negative functions f, g we have

$$\mathbb{E}[f(X)^\alpha g(Y)^\beta] \le \mathbb{E}[f(X)]^\alpha \mathbb{E}[g(Y)]^\beta.$$

Now let us give a direct proof of this result.

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{F}_t)$ be a filtration carrying an $\mathbb{R}^{2n}$–valued Brownian motion $(B_t, \widetilde{B}_t)$ starting from 0 and having covariation given by

$$[(B, \widetilde{B})]_t = t \begin{pmatrix} I_n & \rho\, I_n \\ \rho I_n & I_n \end{pmatrix}.$$

Then $(B_1, \widetilde{B}_1) = (X, Y)$ in law. Assume (without loss of generality) that f and g are bounded away from 0 and $+\infty$ and consider the martingales $(M_t)_{t\in[0,1]}$ and $(N_t)_{t\in[0,1]}$ given by

$$M_t = \mathbb{E}[f(B_1) \mid \mathcal{F}_t], \quad N_t = \mathbb{E}[g(\widetilde{B}_1) \mid \mathcal{F}_t].$$

These are square integrable Brownian martingales, so there exist two $\mathbb{R}^n$–valued processes (u_t) and (v_t) satisfying

$$\mathbb{E}\left[\int_0^1 |u_s|^2 \mathrm{d}s\right] < +\infty, \quad \mathbb{E}\left[\int_0^1 |v_s|^2 \mathrm{d}s\right] < +\infty \tag{2.1}$$

and such that

$$dM_t = \langle u_t, dB_t \rangle, \quad dN_t = \langle v_t, d\widetilde{B}_t \rangle.$$

From the covariation structure of $(B, \widetilde{B})$ we obtain the following expression for the covariation of the process (M_t, N_t):

$$d[M]_t = |u_t|^2\, \mathrm{d}t, \quad d[N]_t = |v_t|^2\, \mathrm{d}t, \quad d[M, N]_t = \rho\, \langle u_t, v_t \rangle\, \mathrm{d}t.$$

Then by Itô's formula (omitting the variables (M_t, N_t) in the right hand side) we have

$$\begin{aligned} d\psi(M_t, N_t) &= \partial_x \psi\, \langle u_t, dB_t \rangle + \partial_y \psi\, \langle v_t, d\widetilde{B}_t \rangle \\ &\quad + \frac{1}{2}\left(\partial^2_{xx}\psi\, |u_t|^2 + \partial^2_{yy}\psi\, |v_t|^2 + 2\partial^2_{xy}\psi\, \rho\, \langle u_t, v_t \rangle\right) \mathrm{d}t. \end{aligned}$$

Applying this to the function $\psi(x,y) = x^\alpha y^\beta$ we get

$$\begin{aligned} dM_t^\alpha N_t^\alpha &= M_t^\alpha N_t^\alpha \left(\alpha \langle \widetilde{u}_t, dB_t\rangle + \beta \langle \widetilde{v}_t, d\widetilde{B}_t\rangle\right) \\ &\quad + \frac{1}{2} M_t^\alpha N_t^\alpha \left(\alpha(\alpha-1)\, |\widetilde{u}_t|^2 + \beta(\beta-1)\, |\widetilde{v}_t|^2 + 2\alpha\beta\rho \,\langle \widetilde{u}_t, \widetilde{v}_t\rangle\right) \mathrm{d}t, \end{aligned} \tag{2.2}$$

where $\widetilde{u}_t = u_t/M_t$ and $\widetilde{v}_t = v_t/N_t$. Recall that the processes (M_t) and (N_t) are bounded away from 0 and $+\infty$ and recall (2.1). This guarantees that the local martingale part (2.2) is a genuine martingale. Now consider the 2×2 matrix

$$A = \begin{pmatrix} \alpha(\alpha-1) & \alpha\rho \\ \beta\rho & \beta(\beta-1) \end{pmatrix}.$$

Since α and β belong to $[0,1]$ the diagonal coefficients are non–positive. Moreover the hypothesis $\rho^2 \le (1-1/\alpha)(1-1/\beta)$ shows that its determinant is non–negative. So A is a negative matrix. This shows that the absolutely continuous part of (2.2) is non–positive. Therefore $(M_t^\alpha N_t^\beta)$ is a super–martingale, in particular

$$\mathbb{E}[M_1^\alpha N_1^\beta] \le M_0^\alpha N_0^\beta,$$

which is the result. □

Remarks (Reversed hypercontractivity)**.** *As we have seen before, hypercontractivity and log–Sobolev are equivalent, so this provides an alternative proof of the logarithmic Sobolev inequality. The same proof shows that if $\rho^2 \le (1-1/\alpha)(1-1/\beta)$ and α and β are both larger than* 1 *then the inequality is reversed:*

$$\mathbb{E}[f(X)^\alpha g(Y)^\beta)] \ge \mathbb{E}[f(x)]^\alpha \mathbb{E}[g(Y)]^\beta,$$

for every non–negative f, g. In terms of the semigroup (P_t) this can be reformulated as

$$\|P_t f\|_{p(t)} \ge \|f\|_p,$$

for every $f \ge 0$, for every $p \in [0,1]$, where $p(t) = 1 + (p-1)\mathrm{e}^{2t}$. This inequality is called reversed hypercontractivity. *Beware that $\|f\|_p = (\int_{\mathbb{R}^n} |f|^p \, d\gamma_n)^{1/p}$ is no longer a norm, and that as opposed to the direct one, the reversed inequality is only valid for non–negative functions f.*

2.2 Spectral gap and Hermite polynomials

We first focus on the one dimensional case $n = 1$.

Lemma 2.3 (Density)**.** *The set of polynomials $\mathbb{R}[X]$ is dense in $\mathbb{L}^2(\gamma_1)$.*

Proof. For any $f \in \mathbb{L}^2(\gamma_1)$, the Laplace transform T_μ of the signed measure $\mu(dx) = f(x)\gamma_1(dx)$ is finite on $\mathbb{R}$ since for any $\theta \in \mathbb{R}$, by Cauchy–Schwarz inequality,

$$(T_\mu(\theta))^2 = \left(\int \mathrm{e}^{\theta x}\, \mu(dx)\right)^2 \le \int f^2 \,\mathrm{d}\gamma_1 \int \mathrm{e}^{2\theta x}\, \gamma_1(dx) < +\infty.$$

In particular T_μ is analytic on a neighborhood of the origin. As a consequence, if $f \perp \mathbb{R}[X]$ in $\mathbb{L}^2(\mathbb{R})$, then the derivatives of arbitrary order of T_μ vanish at the origin, and therefore T_μ is identically zero. Hence $f = 0$ in $\mathbb{L}^2(\gamma_1)$. □

Definition 2.4 (Hermite polynomials)**.** The Hermite polynomials $(H_k)_{k\geq 0}$ are the orthogonal polynomials obtained from the canonical basis of $\mathbb{R}[X]$ by using the Gram–Schmidt algorithm in $\mathbb{L}^2(\gamma_1)$. They are normalized in such a way that the coefficient of the term of highest degree in H_k is 1 for any $k \geq 0$.

We find $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, ... The density of $\mathbb{R}[X]$ in $\mathbb{L}^2(\gamma_1)$ means that $(H_k)_{k\geq 0}$ is a complete orthogonal system in the Hilbert space $\mathbb{L}^2(\gamma_1)$.

Lemma 2.5 (Hermite polynomials)**.** *Hermite polynomials $(H_k)_{k\geq 0}$ satisfy...*

- *Generating series: for any $k \geq 0$ and $x \in \mathbb{R}$,*

$$H_k(x) = \partial_1^k G(0,x) \quad \textit{where} \quad G(s,x) = \mathrm{e}^{sx-\frac{1}{2}s^2} = \sum_{k=0}^{\infty} \frac{s^k}{k!} H_k(x);$$

- *Three terms recursion formula: for any $k \geq 0$ and $x \in \mathbb{R}$,*

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x);$$

- *Recursive differential equation: for any $k \geq 0$ and $x \in \mathbb{R}$,*

$$H_k'(x) = kH_{k-1}(x);$$

- *Differential equation: for any $k \geq 0$ and $x \in \mathbb{R}$,*

$$H_k''(x) - xH_k'(x) + kH_k(x) = 0.$$

- *The sequence $(H_k/\sqrt{k!})_{k\geq 0}$ is orthonormal in $\mathbb{L}^2(\gamma_1)$ and its span is dense.*

Proof. Exercise! Let us prove the last statement. By definition, Hermite polynomials are orthogonal in $\mathbb{L}^2(\gamma_n)$, the span of $(H_k)_{k\geq 0}$ is the set of polynomials $\mathbb{R}[X]$, we already know that this set it is dense in $\mathbb{L}^2(\gamma_n)$, and by Plancherel's formula,

$$\sum_{k\geq 0} \frac{s^{2k}}{k!^2} \|H_k\|_2^2 = \int G(s,x)^2 \, \gamma_1(dx) = \mathrm{e}^{-s^2} \int \mathrm{e}^{2sx} \, \gamma_1(dx) = \mathrm{e}^{s^2} = \sum_{k\geq 0} \frac{s^{2k}}{k!},$$

which gives $\|H_k\|_2^2 = k!$ by identifying the series coefficients. $\square$

Theorem 2.6 (Hermite polynomials and Ornstein–Uhlenbeck process)**.** *For any $k \geq 0$ and $t \geq 0$, the polynomial H_k is an eigenvector of the O.-U. semigroup P_t (respectively O.-U. infinitesimal generator L) associated to the eigenvalue e^{-kt} (respectively $-k$), in other words, for any $f = \sum_{k\geq 0} a_k H_k \in \mathbb{L}^2(\gamma_1)$ with $k!a_k = \langle f, H_k \rangle$ and for any $t \geq 0$,*

$$Lf = -\sum_{k\geq 1} ka_k H_k \quad \textit{and} \quad P_t f = \sum_{k\geq 0} \mathrm{e}^{-kt} a_k H_k, \quad t \geq 0.$$

This provides a quick proof of the Poincaré inequality, namely, if $f \in \mathbb{L}^2(\gamma_1)$,

$$a_0 = \int f H_0 \mathrm{d}\gamma_1 = \int f \, \mathrm{d}\gamma_1 \quad \text{and} \quad \mathrm{Var}_{\gamma_1}(f) = \sum_{k\geq 1} k! a_k^2 \leq \sum_{k\geq 1} kk! a_k^2 = -\int f L f \, \mathrm{d}\gamma_1,$$

where equality is achieved if $a_k = 0$ for $k > 1$, in other words if $f(x) = H_1(x) = x$. We can also deduce a quick proof of the exponential decay of the variance along the semigroup (which is in fact equivalent to the Poincaré inequality), namely, with the same notations, for any $f \in \mathbb{L}^2(\gamma_1)$, the invariance of γ_1 gives

$$a_0 = \langle f, H_0 \rangle = \int f \, \mathrm{d}\gamma_1 = \int P_t f \, \mathrm{d}\gamma_1,$$

and therefore, for any $t \geq 0$,

$$\mathrm{Var}_{\gamma_n}(P_t f) = \|P_t f - a_0\|_2^2 = \sum_{k\geq 1} a_k^2 \mathrm{e}^{-2kt} k! \leq \mathrm{e}^{-2t} \sum_{k\geq 1} a_k^2 k! = \mathrm{e}^{-2t} \|f - a_0\|_2^2.$$

The gap between the first eigenvalue 0 and the second eigenvalue -1 of L is of length 1. This spectral gap produces the exponential convergence. More generally, the semigroup preserves the spectral decomposition. If $f \perp \mathrm{span}\{H_1, \ldots, H_{k-1}\}$ in $\mathbb{L}^2(\gamma_1)$ then $P_t(f) \perp \mathrm{span}\{H_1, \ldots, H_{k-1}\}$ for any $t \geq 0$ and

$$\|P_t f - \gamma_1(f)\|_2 \leq \mathrm{e}^{-kt} \|f - \gamma_1(f)\|_2.$$

Proof of Theorem 2.6. The ordinary differential equation satisfied by Hermite polynomials rewrites $LH_k = -kH_k$, which means that H_k is also an eigenvector of L for the eigenvalue $-k$. It remains to shows that H_k is an eigenvector of P_t for the eigenvalue e^{-kt} for any $t \geq 0$. To see it, set $s, t \in \mathbb{R}$ and $x \in \mathbb{R}$, and let $Z \sim \gamma_1$, then

$$P_t(G(s,\cdot))(x) = \mathrm{e}^{s\mathrm{e}^{-t}x - \frac{1}{2}s^2} \mathbb{E}\big(\mathrm{e}^{s\sqrt{1-\mathrm{e}^{-2t}}Z}\big)$$

Since the Laplace transform of Z is given by $\mathbb{E}(\mathrm{e}^{\theta Y}) = \mathrm{e}^{\frac{1}{2}\theta^2}$ we get

$$P_t(G(s,\cdot))(x) = G(s\mathrm{e}^{-t}, x).$$

Now, using the generating series property of Hermite polynomials, we get

$$\begin{aligned} P_t(H_k)(x) &= P_t(\partial_1^k G(0,\cdot))(x) = \partial_s^k P_t(G(s,\cdot))(x)_{|s=0} \\ &= \partial_s^k G(s\mathrm{e}^{-t}, x)_{|s=0} = \mathrm{e}^{-kt} \partial_1^k G(s\mathrm{e}^{-t}, x)_{|s=0} = \mathrm{e}^{-kt} H_k(x). \end{aligned}$$

□

The same holds for the Ornstein–Uhlenbeck process in any dimension $n \geq 1$ with the products of Hermite polynomials. First it can be checked that the set of n-variate polynomials $\mathbb{R}[X_1, \ldots, X_n]$ is dense in $\mathbb{L}^2(\gamma_n)$, and that the tensor products of Hermite polynomials $(H_{k_1,\ldots,k_n})_{k_1,\ldots,k_n \in \mathbb{N}}$ where

$$H_{k_1,\ldots,k_n}(X_1, \ldots, X_n) = H_{k_1}(X_1) \cdots H_{k_n}(X_n)$$

form an orthogonal family in $\mathbb{L}^2(\gamma_n)$. The O.–U. infinitesimal generator in $\mathbb{R}^n$ writes

$$L = \Delta - \langle x, \nabla \rangle = L_1 + \cdots + L_n$$

where $L_i = \partial_i^2 - x_i \partial_i$ is the one-dimensional Ornstein–Uhlenbeck operator acting on the i-th variable, and therefore, for any $k_1, \ldots, k_n \in \mathbb{N}$,

$$L(H_{k_1,\ldots,k_n}) = -(k_1 + \cdots + k_n) H_{k_1,\ldots,k_n}.$$

The spectral gap of L is equal to -1 for any dimension $n \geq 1$, while the associated eigenspace is $\mathrm{span}(H_1(x_1), \ldots, H_1(x_n))$ and is of dimension is n.

Theorem 2.7 (Quantum harmonic oscilliator)**.** *If we define $\Theta : \mathbb{L}^2(\mathrm{d}x) \to \mathbb{L}^2(\gamma_n)$ by*

$$\Theta(f) = f\varphi_n^{-1/2}$$

where φ_n is the Lebesgue density of γ_n, then Θ is a linear isometry, the operator

$$\widetilde{L} := \Theta^{-1} \circ L \circ \Theta$$

satisfies for any smooth enough $f \in \mathbb{L}^2(\mathrm{d}x)$ the formula

$$\widetilde{L}f(x) = \varphi_n^{1/2} L(f\varphi_n^{-1/2}) = \Delta f + \left(\frac{n}{2} - \frac{|x|^2}{4}\right) f,$$

and for any $(k_1, \ldots, k_n) \in \mathbb{N}^n$ the function

$$\Psi_{k_1,\ldots,k_n}(x) = \Theta^{-1}(H_{k_1,\ldots,k_n})(x_1, \ldots, x_n) = \mathrm{e}^{-\frac{1}{4}|x|^2} H_{k_1}(x_1) \cdots H_{k_n}(x_n), \quad x \in \mathbb{R}^n,$$

is an eigenvector of $\widetilde{L}$, namely

$$\widetilde{L}\Psi_{k_1,\ldots,k_n} = -(k_1 + \cdots + k_n)\Psi_{k_1,\ldots,k_n}.$$

The operator $\widetilde{L}$ is a Schrödinger operator: the sum of a Laplacian with a multiplicative potential. It is known as the “quantum harmonic oscilliator”. In the quantum mechanics modelling, the eigenvectors of $\widetilde{L}$ are “wave functions”, and it turns out here that they are nothing else but Hermite polynomials damped by a Gaussian weight.

Proof. The fact that Θ is an isometry follows from

$$\int (\Theta(f))^2 \, \mathrm{d}\gamma_n = \int f^2 \varphi_n^{-1} \varphi_n \, \mathrm{d}\gamma_n = \int f^2 \, \mathrm{d}x.$$

Since L and $\widetilde{L}$ are isometric, they share the same spectrum and their eigenvectors are in bijection: the ones of $\widetilde{L}$ are the image of the ones of L by Θ^{-1}.

The explicit formula for $\widetilde{L}$ can be obtained using the “algebraic” formulas

$$L(\beta(f)) = \beta'(f)Lf + \beta''(f)|\nabla f|^2 \quad \text{and} \quad L(fg) = fLg + gLf + 2\nabla f \cdot \nabla g.$$

and $\varphi_n = \mathrm{e}^{-V}$ and $L = \Delta - \langle \nabla V, \nabla \rangle$ with $V(x) = (1/2)|x|^2 + (n/2)\log(2\pi)$. □

Remark (Beyond the Ornstein–Uhlenbeck exactly solvable model)**.** *More generally, if one has the Markov diffusion process $X = (X_t)_{t\geq 0}$ on $\mathbb{R}^n$ solving the stochastic differential equation $\mathrm{d}X_t = \sqrt{2}\mathrm{d}B_t - \nabla U(X_t)\mathrm{d}t$ where $B = (X_t)_{t\geq 0}$ is a standard Brownian motion, with infinitesimal generator $L = \Delta - \nabla U \cdot \nabla$ and reversible invariant probability measure $\mu(\mathrm{d}x) = \mathrm{e}^{-U(x)}\mathrm{d}x$, and if one considers the isometry $\Theta : \mathbb{L}^2(\mathrm{d}x) \to \mathbb{L}^2(\mu)$ defined by $\Theta(f) = \mathrm{e}^{\frac{1}{2}V} f$ then the conjugated operator $\widetilde{L} = \Theta^{-1} \circ L \circ \Theta$ is the Schrödinger operator*

$$\widetilde{L}(f)(x) = \Delta f(x) + U(x)f(x) \quad \textit{where} \quad V = \frac{\Delta V}{2} - \frac{|\nabla V|^2}{4}.$$

This formula appears in the Girsanov formula giving the density of the law of the sample paths of X with respect to the law of the driving Brownian motion B, see [28, 27]. The eigenvalues and eigenvectors are known for very special exactly solvable cases only such as the Ornstein–Uhlenbeck case associated to the quadratic potential $U(x) = \frac{1}{2}|x|^2 + \frac{n\log(2\pi)}{2}$.

For instance, for $(k_1,\dots,k_n)=(0,1,\dots,n-1)$, we get the wave function

$$\psi(x_1,\dots,x_n)=\sqrt{\varphi_n(x)}H_0(x_1)\cdots H_{n-1}(x_n).$$

In mathematical physics, a bosonic wave function of n particles is obtained by symmetrization over $x_1,\dots,x_n$. A fermionic wave function is obtained by anti-symmetrization, which implies nullity on the diagonal, for instance

$$\begin{aligned}\psi_{\text{fermions}}(x_1,\dots,x_n)&=\sqrt{\varphi_n(x)}\sum_{\sigma\in\Sigma_n}(-1)^{\text{signature}(\sigma)}H_{\sigma(1)-1}(x_1)\cdots H_{\sigma(n)-1}(x_n)\\&=\sqrt{\varphi_n(x)}\det\begin{pmatrix}H_0(x_1)&\dots&H_0(x_n)\\\vdots&\vdots&\vdots\\H_{n-1}(x_1)&\dots&H_{n-1}(x_n)\end{pmatrix}\\&=\sqrt{\varphi_n(x)}\det\begin{pmatrix}x_1^0&\dots&x_n^0\\\vdots&\vdots&\vdots\\x_1^{n-1}&\dots&x_n^{n-1}\end{pmatrix}\\&=\sqrt{\varphi_n(x)}\prod_{1\le i<j\le n}(x_i-x_j).\end{aligned}$$

This "Slater determinant" is proportional to a Vandermonde determinant. It is remarkable to see that $(x_1,\dots,x_n)\in\mathbb{R}^n\mapsto \mathrm{e}^{-\frac14|x|^2}\prod_{1\le i<j\le n}(x_i-x_j)$ is an eigenvector of $\widetilde{L}$. Now

$$|\psi_{\text{fermions}}(x_1,\dots,x_n)|^2=(2\pi)^{-\frac{n}{2}}\mathrm{e}^{-\frac12(x_1^2+\cdots+x_n^2)}\prod_{1\le i<j\le n}(x_i-x_j)^2.$$

We recognize, up to normalization, the formula of the probability density function of the famous "Gaussian Unitary Ensemble" (GUE) namely the probability density function of the eigenvalues of a Gaussian $n\times n$ Hermitian random matrix with Lebesgue density in $\mathbb{R}^{n+n^2-n}=\mathbb{R}^{n^2}$ proportional to $H\mapsto\exp(-\frac12\mathrm{Tr}(H^2))$.

2.3 Information theory

The Boltzmann entropy was introduced in statistical physics by Ludwig Boltzmann. It was introduced latter by Claude Shannon in information theory.

Definition 2.8 (Boltzmann or Shannon entropy)**.** Let μ be a probability measure on $\mathbb{R}^n$ with Lebesgue density f such that $f\log f$ is Lebesgue integrable. The Boltzmann or Shannon entropy of μ is defined by

$$\mathrm{S}(X)=-\int_{\mathbb{R}^n}f\log f\,dx.$$

In other words $-\mathrm{S}(\mu)$ is the relative entropy of μ with respect to the Lebesgue measure, but in the sequel the relative entropy functional H is used exclusively between probability measures. Still about notations, when X is a random vector on $\mathbb{R}^n$ of law μ, we write $\mathrm{S}(X)$ for $\mathrm{S}(\mu)$. We define $\mathrm{H}(X\mid\gamma_n)$ similarly.

Remark (Relative entropy or free energy?)**.** *A standard object of statistical physics is a probability measure $\mu=\mu_{\beta,V}$ on say $\mathbb{R}^n$ with Lebesgue density of the form $x\in\mathbb{R}^n\mapsto Z^{-1}\mathrm{e}^{-\beta V(x)}$ where $V(x)$ is the "energy" of the configuration $x\in\mathbb{R}^n$, the parameter $\beta>0$*

is an "inverse temperature", and $Z = Z_{\beta,V}$ is the normalizing constant. If ν is another probability measure with Lebesgue density h, then

$$\mathrm{H}(\nu \mid \mu) = \int h \log h \, dx + \beta \int V \, \mathrm{d}\nu + \log Z = \beta \int V \, \mathrm{d}\nu - S(\nu) + \log Z.$$

In thermodynamics $A(\nu) := \int V \mathrm{d}\nu - \frac{1}{\beta} S(\nu)$ is a Helmholtz "free energy", and

$$A(\mu) = -\frac{1}{\beta} \log Z \quad \textit{and} \quad A(\nu) - A(\mu) = \frac{1}{\beta} H(\nu \mid \mu) \geq 0.$$

From the variational point of view, the maximization of S under a mean energy constraint leads to the Boltzmann–Gibbs measure $\mu_{\beta,V}$, and this constrained optimization problem is equivalent to the minimization of A without constraint, in other words A is a Lagrangian.

The Shannon entropy is translation invariant: $\mathrm{S}(X+m) = \mathrm{S}(X)$ for any random vector X and any $m \in \mathbb{R}$. It also has the following scaling property

$$\mathrm{S}(\lambda X) = S(X) + n \log(|\lambda|)$$

for every $\lambda \neq 0$. More generally, for every invertible $n \times n$ matrix A we have

$$\mathrm{S}(AX) = \mathrm{S}(X) + \log(|\det(A)|)$$

The proof is left as an exercise.

The following theorem states that among vectors having fixed invertible covariance matrix K, the Gaussian vector maximizes the entropy.

Theorem 2.9 (Gaussian maximum entropy). *If X has finite entropy and covariance matrix K and if G is a Gaussian vector with covariance K then*

$$\mathrm{S}(X) \leq \frac{n}{2} \log(2\pi \mathrm{e} (\det K)^{1/n}) = \mathrm{S}(G).$$

Moreover there is equality if and only if X is Gaussian.

Proof. Since the Shannon entropy is translation invariant one can assume that X and G have zero mean. Let f and g be their respective densities, in particular

$$g(x) = ((2\pi)^n \det K)^{-1/2} \, \mathrm{e}^{-\langle x, K^{-1}x \rangle/2}.$$

Since X and G are centered with the same covariance matrix, and since $\log g$ is a quadratic form we have $\mathbb{E}[\log g(X)] = \mathbb{E}[\log g(G)]$. Therefore

$$\mathbb{E}[\log f(X)] = \mathbb{E}[\log(f/g)(X)] + \mathbb{E}[\log g(X)] = \mathbb{E}[\log(f/g)(X)] + \mathbb{E}[\log g(G)].$$

In other words $\mathrm{S}(G) - \mathrm{S}(X) = \mathrm{H}(X \mid G)$. Since the relative entropy $\mathrm{H}(X \mid G)$ is non negative and vanishes only when $X = G$ in law, this is the result. □

Theorem 2.10 (Sub-additivity). *If $X = (X_1, \ldots, X_n)$ is a randon vector of $\mathbb{R}^n$ with finite entropy then*

$$S(X_1, \ldots, X_n) \leq S(X_1) + \cdots + S(X_n)$$

with equality if and only X has independent components $X_1, \ldots, X_n$.

Proof. Let f be the density of X, and $f_i = \int f \prod_{j \neq i} dx_j$ be the density of X_i. Let μ and ν be the probability measures with densities f and $f_1 \otimes \cdots \otimes f_n$. Then

$$S(X_1) + \cdots + S(X_n) - S(X_1, \ldots, X_n) = \mathrm{H}(\mu \mid \nu) \geq 0.$$

Moreover equality is achieved if and only if $\mu = \nu$. □

Here is the main result of this section

Theorem 2.11 (Shannon–Stam inequality). *Let X, Y be independent random vectors on $\mathbb{R}^n$ and let $\theta \in [0, 1]$. Then*

$$\mathrm{S}\left(\sqrt{1-\theta}X + \sqrt{\theta}Y\right) \geq (1-\theta)\mathrm{S}(X) + \theta\mathrm{S}(Y)$$

Using the scaling properties of the Shannon entropy one can sharpen the previous result as follows. Given a random vector X we let

$$\mathrm{N}(X) = \frac{1}{2\pi \mathrm{e}} \mathrm{e}^{\frac{2}{n}\mathrm{S}(X)}.$$

This is called the (Shannon) *entropy power* of X. The normalization insures that if G is a standard Gaussian vector then $\mathrm{N}(G) = 1$. If $X \sim \mathcal{N}(0, K)$ then $N(X) = \det(K)^{1/n}$. The scaling properties of the entropy show that the entropy power is translation invariant and 2–homogeneous in the sense that $\mathrm{N}(\lambda X + m) = \lambda^2 \mathrm{N}(X)$.

Corollary 2.12 (Entropy power inequality). *If X, Y are independent random vectors of the same dimension then*

$$\mathrm{N}(X + Y) \geq \mathrm{N}(X) + \mathrm{N}(Y).$$

Proof. Let $\theta \in [0, 1]$. Using Shannon–Stam and the scaling property of the entropy we get

$$\begin{aligned} \mathrm{S}(X+Y) &= \mathrm{S}\left(\sqrt{1-\theta}\frac{X}{\sqrt{1-\theta}} + \sqrt{\theta}\frac{Y}{\sqrt{\theta}}\right) \\ &\geq (1-\theta)\left(\mathrm{S}(X) - \frac{n}{2}\log(1-\theta)\right) + \theta\left(\mathrm{S}(Y) - \frac{n}{2}\log\theta\right). \end{aligned}$$

In other words

$$\mathrm{N}(X+Y) \geq \left(\frac{\mathrm{N}(X)}{1-\theta}\right)^{1-\theta} \left(\frac{\mathrm{N}(Y)}{\theta}\right)^{\theta}.$$

Then choose $\theta = \mathrm{N}(Y)/(\mathrm{N}(X) + \mathrm{N}(Y))$. □

The purpose of the rest of this section is twofold:

1. Give a proof of the Shannon–Stam inequality;

2. Draw a connection with the logarithmic Sobolev inequality.

Definition 2.13 (Fisher information). Let μ be a probability measure on $\mathbb{R}^n$ having a smooth density f with respect to the Lebesgue measure. The *Fisher information* of μ, introduced by Ronald Fisher in mathematical statistics, is

$$\mathrm{J}(\mu) = \int_{\mathbb{R}^n} |\nabla \log f|^2 \, \mathrm{d}\mu.$$

More generally, the relative Fisher information of μ with respect to ν is

$$\mathrm{J}(\mu \mid \nu) = \int_{\mathbb{R}^n} |\nabla \log(\mathrm{d}\mu/d\nu)|^2 \, \mathrm{d}\mu,$$

provided μ has a smooth density with respect to ν.

Again when X is a random vector we let $\mathrm{J}(X)$ be the Fisher information of the law of X. The Fisher information is translation invariant and -2–homogeneous: for every random vector X and every $\lambda \neq 0$ and $m \in \mathbb{R}^n$ we have

$$\mathrm{J}(\lambda X + m) = \lambda^{-2}\,\mathrm{J}(X).$$

Definition 2.14. Let X be a random vector in $\mathbb{R}^n$ whose law has smooth density f with respect to the Lebesgue measure. The *score* of X is the random vector $\rho_X = \nabla \log f(X)$. In particular we have

$$\mathbb{E}[|\rho_X|^2] = \mathrm{J}(X).$$

Note the score has the following homogeneity property: $\rho_{\lambda X} = \lambda^{-1}\rho_X$. Also, integrating by parts we get

$$\mathbb{E}[\rho_X] = \int_{\mathbb{R}^n} (\nabla \log f) f \, dx = \int_{\mathbb{R}^n} \nabla f \, dx = 0,$$

for every vector X.

Lemma 2.15. *Let X and Y be independent random vector having smooth densities with respect to the Lebesgue measure. If the score of X is integrable then*

$$\rho_{X+Y} = \mathbb{E}[\rho_X \mid X+Y].$$

Proof. Let f and g be the respective densities of X and Y and let h be a bounded and measurable function on $\mathbb{R}^n$. Then we write

$$\begin{aligned}
\mathbb{E}[\rho_X\, h(X+Y)] &= \int_{\mathbb{R}^{2n}} \nabla \log f(x) h(x+y) f(x) g(y)\, dxdy \\
&= \int_{\mathbb{R}^n} (\nabla f) * g(z)\, h(z)\, dz \\
&= \int_{\mathbb{R}^n} \nabla (f*g)(z) h(z)\, dz = \mathbb{E}[\rho_{X+Y} h(X+Y)],
\end{aligned}$$

which is the result. □

Since conditional expectation contract the $\mathbb{L}^2$–norm, it follows immediately from the previous lemma that if X and Y are independent then $\mathrm{J}(X+Y) \leq \mathrm{J}(X)$. With a little extra work we can actually get the following

Theorem 2.16 (Blachmann–Stam inequality)**.** *If X, Y are independent random vectors of the same dimension and $\theta \in [0,1]$ then then*

$$\mathrm{J}\left(\sqrt{1-\theta}X + \sqrt{\theta}Y\right) \leq (1-\theta)\mathrm{J}(X) + \theta \mathrm{J}(Y).$$

Proof. Set $Z = \sqrt{1-\theta}X + \sqrt{\theta}Y$. Lemma 2.15 and the homogeneity of the score give

$$\rho_Z = \mathbb{E}[(1-\theta)^{-1/2}\rho_X \mid Z]$$

and similarly $\rho_Z = \mathbb{E}[\theta^{-1/2}\rho_Y \mid Z]$. Multiplying the first equality by $1-\theta$, the second one by θ and adding them together we get

$$\rho_Z = \mathbb{E}[\sqrt{1-\theta}\,\rho_X + \sqrt{\theta}\,\rho_Y \mid Z].$$

Taking the norm squared and expectation (and using Jensen) we get

$$\mathrm{J}(Z) \leq \mathbb{E}[|\sqrt{1-\theta}\rho_X + \sqrt{\theta}\rho_Y|^2].$$

Now ρ_X and ρ_Y are independent and centered, so if we expand the square, the cross term vanishes and we get the result. □

We have seen in previous sections that the derivative of the relative entropy along the Ornstein–Uhlenbeck semigroup is minus the relative Fisher information:

$$\frac{d}{\mathrm{d}t}\mathrm{H}(X_t \mid \gamma_n) = -\int \frac{|\nabla g_t|^2}{g_t}\,\mathrm{d}\gamma_n = -\mathrm{J}(X_t \mid \gamma_n)$$

where g_t is the density of X_t with respect to γ_n. Here is a similar formula for the Shannon entropy along the heat semigroup.

Theorem 2.17 (de Bruijn identity)**.** *Let X be a random vector with finite entropy and G be a standard Gaussian random vector independent of X. Then*

$$\frac{d}{\mathrm{d}t}\mathrm{S}(X+\sqrt{t}G) = \frac{1}{2}\mathrm{J}(X+\sqrt{t}G).$$

Proof. Let f_t be the density of $X+\sqrt{t}Z$. It can be checked that $(f_t)_{t\geq 0}$ solves the heat equation: $\partial_t f_t = (1/2)\Delta f_t$. Now, using an integration by parts,

$$\begin{aligned}\frac{d}{\mathrm{d}t}\mathrm{S}(X+\sqrt{t}Z) &= -\frac{d}{\mathrm{d}t}\int f_t \log f_t \,dx\\ &= -\int \frac{1}{2}(\Delta f_t)(1+\log f_t)\,dx\\ &= \frac{1}{2}\int \langle \nabla f_t \cdot \nabla \log f_t\rangle \,dx\\ &= \frac{1}{2}J(X+\sqrt{t}Z).\end{aligned}$$

The heat semigroup $P := (P_t)_{t\geq 0}$ satisfies $P_t(f) = f * \gamma_t$ where γ_t is the density of tZ. The Lebesgue measure μ on $\mathbb{R}^n$ is invariant and symmetric for P and

$$\mathrm{H}(X+\sqrt{t}Z) = -\int \phi(P_t(f_0)))\,\mathrm{d}\mu.$$

The function $P_t(f_0) = f_0 * \gamma_t$ is the density of B_t^X where $(B_t^X)_{t\geq 0}$ is a standard Brownian motion started from the random initial condition X of density f_0. □

We are now in a position to prove the Shannon–Stam inequality. The proof combines Blachmann–Stam with the de Bruijn identity.

Proof of Shannon–Stam. Let X and Y be independent random vectors having finite entropy and let $G, \widetilde{G}$ be two independent standard Gaussian vectors, independent of X and Y. Let $X_t = \sqrt{1-t}X + \sqrt{t}G$. Writing

$$\mathrm{S}(X_t) = \mathrm{S}\left(X + \frac{\sqrt{t}}{\sqrt{1-t}}G\right) + \frac{n}{2}\log(1-t)$$

and using the de Bruijn identity we get

$$\frac{d}{\mathrm{d}t}\mathrm{S}(X_t) = \frac{1}{2(1-t)}\left(\mathrm{J}(X_t) - n\right).$$

Therefore

$$\mathrm{S}(G) - \mathrm{S}(X) = \mathrm{S}(X_1) - \mathrm{S}(X_0) = \frac{1}{2}\int_0^1 \frac{\mathrm{J}(X_t)-n}{1-t}\,\mathrm{d}t.$$

Of course we have a similar equality for $Y_t = \sqrt{1-t}\,Y + \sqrt{t}\,\widetilde{G}$. Now we fix $\theta \in [0,1]$, we let $Z = \sqrt{1-\theta}\,X + \sqrt{\theta}\,Y$ and, more generally, $Z_t = \sqrt{1-\theta}\,X_t + \sqrt{\theta}\,Y_t$ for all $t \in [0,1]$. Since X_t and Y_t are independent the Blachmann–Stam inequality gives

$$\mathrm{J}(Z_t) \leq (1-\theta)\,\mathrm{J}(X_t) + \theta\,\mathrm{J}(Y_t) \tag{2.3}$$

for all t. On the other, observe that $Z_t = \sqrt{1-t}\,Z + \sqrt{t}\,W$ where $W = \sqrt{1-\theta}\,G + \sqrt{\theta}\,\widetilde{G}$ is a standard Gaussian independent of Z. So we also have

$$\mathrm{S}(G) - \mathrm{S}(Z) = \frac{1}{2}\int_0^1 \frac{\mathrm{J}(Z_t) - n}{1-t}\,\mathrm{d}t.$$

So substracting n from (2.3) dividing by $1-t$ and integrating between 0 and 1 we get $\mathrm{S}(Z) \geq (1-\theta)\mathrm{S}(X) + \theta\mathrm{S}(Y)$ which is the result. □

Now we draw a connection with the log–Sobolev inequality. Let X be a random vector and let G be a standard Gaussian vector independent of X. Applying the Entropy power inequality to X and $\sqrt{t}\,G$ we get

$$\mathrm{N}(X + \sqrt{t}G) \geq \mathrm{N}(X) + \mathrm{N}(\sqrt{t}G) = \mathrm{N}(X) + t.$$

This shows in particular that

$$\frac{d}{\mathrm{d}t}_{|t=0}\mathrm{N}(X + \sqrt{t}G) \geq 1.$$

On the other hand, using the de Bruijn identity, we easily get

$$\frac{d}{\mathrm{d}t}_{|t=0}\mathrm{N}(X + tG) = \frac{\mathrm{J}(X)\mathrm{N}(X)}{n}.$$

So we have just proved that

$$\mathrm{J}(X)\mathrm{N}(X) \geq n.$$

Observe that the quantity $\mathrm{J}(X)\mathrm{N}(X)$ is scale invariant. The above inequality asserts that it is minimized by Gaussian vectors. Taking the logarithm, we can reformulate this inequality as follows:

Theorem 2.18 (Euclidean form of the logarithmic Sobolev inequality)**.** *For any probability measure μ on $\mathbb{R}^n$ we have*

$$\mathrm{S}(\gamma_n) - \mathrm{S}(\mu) \leq \frac{n}{2}\log\left(\frac{\mathrm{J}(\mu)}{n}\right). \tag{2.4}$$

We close this section by showing the relationship of this with the usual log–Sobolev inequality. The latter asserts that

$$\mathrm{H}(\mu \mid \gamma_n) \leq \frac{1}{2}\mathrm{J}(\mu \mid \gamma_n),$$

for every probability measure μ. Let X be a random vector having law μ. An easy computation shows that

$$\mathrm{H}(X \mid \gamma_n) = -\mathrm{S}(X) + \mathrm{S}(\gamma_n) + \frac{1}{2}\mathbb{E}[|X|^2] - \frac{n}{2}.$$

In the same way, letting f be the density of μ with respect to Lebesgue and g be that of γ_n we have

$$\begin{aligned}\mathrm{J}(X \mid \gamma_n) &= \mathbb{E}[|\nabla \log(f/g)(X)|^2] = \mathbb{E}[|\nabla \log f(X) + X|^2] \\ &= \mathrm{J}(X) + 2\mathbb{E}[\langle \nabla \log f(X), X\rangle] + \mathbb{E}[|X|^2].\end{aligned}$$

Moreover, an integration by parts shows that

$$\mathbb{E}[\langle \nabla \log f(X), X\rangle] = \int_{\mathbb{R}^n} \langle \nabla f(x), x\rangle \, dx = -\int_{\mathbb{R}^n} f(x)\mathrm{div}(x)\, dx = -n.$$

We thus obtain

$$\mathrm{J}(X \mid \gamma_n) = \mathrm{J}(X) + \mathbb{E}[|X|^2] - 2n.$$

Therefore, the log–Sobolev inequality can be reformulated as

$$\mathrm{S}(\gamma_n) - \mathrm{S}(\mu) \leq \frac{1}{2}\mathrm{J}(\mu) - \frac{n}{2}. \tag{2.5}$$

Using the inequality $\log x \leq x - 1$, we immediately see that this is weaker than (2.4). Actually (2.4) can be recovered from its weaker version by a scaling argument (the should not be a surprise, given that EPI was obtained from Shannon the same way). More precisely, applying (2.5) to λX and using the scaling properties of the entropy and the Fisher information, we get

$$\mathrm{S}(\gamma_n) - \mathrm{S}(\mu) \leq \frac{1}{2\lambda^2}\mathrm{J}(\mu) - \frac{n}{2} + n\log\lambda.$$

Now optimizing on λ we obtain back (2.4).

Chapter 3

Sub-Gaussian concentration and transportation

A function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz when

$$\|f\|_{\mathrm{Lip}} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} < \infty.$$

A Lipschitz function is always continuous, and a theorem due to Hans Rademacher states that it is differentiable almost everywhere. When f is $\mathcal{C}^1$ then

$$\|f\|_{\mathrm{Lip}} = \sup_{x \in \mathbb{R}^n} |\nabla f(x)| = \||\nabla f|\|_\infty.$$

3.1 Concentration of measure

Theorem 3.1 (Sub-Gaussian bound on Laplace transform of Lipschitz functions)**.** *For any $n \geq 1$, and any Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, and any $\theta \in \mathbb{R}$,*

$$L(\theta) := \int \exp(\theta f) \mathrm{d}\gamma_n \leq \exp\left(\frac{\theta^2}{2}\|f\|_{\mathrm{Lip}}^2 + \theta \int f \mathrm{d}\gamma_n\right)$$

In other words

$$\int \exp\left(\theta\left(f - \int f \mathrm{d}\gamma_n\right)\right) \mathrm{d}\gamma_n \leq \exp\left(\frac{\theta^2}{2}\|f\|_{\mathrm{Lip}}^2\right)$$

The right hand side does not depend on n.

Proof. First we observe that for any $\theta \in \mathbb{R}$, we have $\mathrm{e}^{\theta f} \in \mathbb{L}^1(\gamma_n)$ since

$$\int \mathrm{e}^{\theta f} \, \mathrm{d}\gamma_n \leq \mathrm{e}^{|\theta||f(0)|} \int \mathrm{e}^{|\theta|\|f\|_{\mathrm{Lip}}|x|} \, \mathrm{d}\gamma_n(x) < \infty.$$

In particular $f \in \mathrm{L}^1(\gamma_n)$. We can assume that f is bounded and $\mathcal{C}^\infty$ by using cutoff, regularization, and Fatou's lemma. Indeed, let f be Lipschitz. Let us define, for all $k \geq 1$ and $\varepsilon > 0$, the function $f_{k,\varepsilon} := \max(-k, \min(f, k)) * \rho_\varepsilon$ where $\rho_\varepsilon \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ satisfies[1]

$$\mathrm{supp}(\rho_\varepsilon) \subset \{x \in \mathbb{R}^n : |x| \leq \varepsilon\}, \quad \rho_\varepsilon \geq 0, \quad \text{and} \int_{\mathbb{R}^n} \rho_\varepsilon(x) \mathrm{d}x = 1.$$

[1]This is a *mollifier*. Example: $\rho_\varepsilon(x) = \varepsilon^{-n}\rho(\varepsilon^{-1}|x|)$ with $\rho(x) := c^{-1}\mathrm{e}^{-\frac{1}{1-x^2}}\mathbf{1}_{|x|<1}$, $c := \displaystyle\int \rho(x)\mathrm{d}x$.

Then $\|f_{k,\varepsilon}\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$ and $f_{k,\varepsilon} \to f$ pointwise and in $\mathbb{L}^1(\gamma_n)$ as $k \to \infty$ and $\varepsilon \to 0$ (see below). If the result is true for $f_{k,\varepsilon}$ then, using the Fatou lemma for the first inequality,

$$\begin{aligned}\int \mathrm{e}^{\theta f}\mathrm{d}\gamma_n = \int \lim_{\substack{k\to\infty\\ \varepsilon\to 0}} \mathrm{e}^{\theta f_{k,\varepsilon}}\mathrm{d}\gamma_n &\leq \varliminf_{\substack{k\to\infty\\ \varepsilon\to 0}} \int \mathrm{e}^{\theta f_{k,\varepsilon}}\mathrm{d}\gamma_n \\ &\leq \varliminf_{\substack{k\to\infty\\ \varepsilon\to 0}} \exp\Big(\frac{\theta^2}{2}\|f_{k,\varepsilon}\|^2_{\mathrm{Lip}} + \theta\int f_{k,\varepsilon}\mathrm{d}\gamma_n\Big) \\ &\leq \exp\Big(\frac{\theta^2}{2}\|f\|^2_{\mathrm{Lip}} + \theta\int f\mathrm{d}\gamma_n\Big).\end{aligned}$$

Thus we can take f smooth and bounded. We may also assume that $\theta > 0$ by replacing f with $-f$, and that f is centered for γ_n and $\|f\|_{\mathrm{Lip}} = 1$ by translation and scaling.

The idea now is as follows:[2] for any $\theta > 0$, the logarithmic Sobolev inequality for the probability measure γ_n and the smooth and bounded test function $\mathrm{e}^{\theta f}$ gives, using $|\nabla \mathrm{e}^{\theta f}| = |\theta\nabla f|\mathrm{e}^{\theta f}$ and $\||\nabla f|\|_\infty = \|f\|_{\mathrm{Lip}} \leq 1$, that

$$\theta L'(\theta) - L(\theta)\log L(\theta) \leq \frac{\theta^2}{2}L(\theta).$$

This can we rewritten as $K' \leq 1/2$ where $K(\theta) := (1/\theta)\log L(\theta)$. Since $L(0) = 1$ and $L'(0) = \gamma_n(f)$, and $K(0) = (\log L)'(0) = L'(0)/L(0) = \gamma_n(f)$, the result follows.

Finally, let us give more details on the approximation argument. We have $f_{k,\varepsilon} = f_k * \rho_\varepsilon$, $f_k := \max(-k, \min(f,k))$, and $\|f_k\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$. Now $\|f_{k,\varepsilon}\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$ since

$$|f_{k,\varepsilon}(x) - f_{k,\varepsilon}(y)| \leq \int_{\mathbb{R}^n} |f_k(x-z) - f_k(y-z)|\rho_\varepsilon(z)\mathrm{d}z \leq \|f\|_{\mathrm{Lip}}|x-y|.$$

For the pointwise convergence of $f_{k,\varepsilon}$ to f, we first note that if $|x-y| \leq \varepsilon$ then

$$|f_k(y) - f(x)| \leq |f_k(y) - f_k(x)| + |f_k(x) - f(x)| \leq \varepsilon\|f\|_{\mathrm{Lip}} + |f(x)|\mathbf{1}_{|f(x)|\geq k} \underset{\substack{k\to\infty\\ \varepsilon\to 0}}{\longrightarrow} 0,$$

and thus, for all $x \in \mathbb{R}^n$,

$$|f_{k,\varepsilon}(x) - f(x)| \leq \int_{|x-y|\leq\varepsilon} |f_k(y) - f(x)|\rho_\varepsilon(x-y)\mathrm{d}y \leq \varepsilon\|f\|_{\mathrm{Lip}} + |f(x)|\mathbf{1}_{|f(x)|\geq k} \underset{\substack{k\to\infty\\ \varepsilon\to 0}}{\longrightarrow} 0.$$

The convergence in $\mathbb{L}^1(\gamma_n)$ of $f_{k,\varepsilon}$ to f follows by dominated convergence. ☐

Corollary 3.2 (Concentration for Lipschitz functions)**.** *For any $n \geq 1$ and any $f : \mathbb{R}^n \to \mathbb{R}$ Lipschitz, and any real $r \geq 0$,*

$$\gamma_n(f \geq \gamma_n(f) + r)) \leq \exp\left(-\frac{r^2}{2\|f\|^2_{\mathrm{Lip}}}\right).$$

The right hand side does not depend on n. Using the result for f and $-f$, we get also

$$\gamma_n(|f - \gamma_n(f)| \geq r) \leq 2\exp\left(-\frac{r^2}{2\|f\|^2_{\mathrm{Lip}}}\right).$$

This means that under γ_n, f is “concentrated” around its mean $\gamma_n(f)$, hence the term. In terms of random variables, the inequality above writes, for any $Z \sim \gamma_n$,

$$\mathbb{P}(|f(Z) - \mathbb{E}(f(Z))| \geq r) \leq 2\exp\left(-\frac{r^2}{2\|f\|^2_{\mathrm{Lip}}}\right).$$

[2]The argument is attributed to Ira Herbst, was communicated to Leonard Gross and Oscar Rothaus [20], and was popularized later on by Michel Ledoux [23].

Proof. We reduce to $\|f\|_{\mathrm{Lip}} = 1$ and $\gamma_n(f) = \int f \,\mathrm{d}\gamma_n = 0$ by scaling and translation. For any $r \geq 0$ and $\theta > 0$, by Markov's inequality and Theorem 3.1,

$$\gamma_n(f \geq r) = \gamma_n\left(\mathrm{e}^{\theta f} \geq \mathrm{e}^{\theta r}\right) \leq \mathrm{e}^{-\theta r}\int \mathrm{e}^{\theta f}\,\mathrm{d}\gamma_n \leq \mathrm{e}^{-\theta r + \frac{1}{2}\theta^2} \leq \mathrm{e}^{-\frac{1}{2}r^2},$$

where the last inequality is obtained by taking the optimal choice $\theta = r$. $\square$

Corollary 3.3 (Quantitative bounds for empirical means)**.** *For any integer $n \geq 1$, if $X_1, \ldots, X_N$ are independent and identically distributed random variables with law γ_n, then for any Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, any integer $N \geq 1$, any real $r \geq 0$,*

$$\mathbb{P}\left(\left|\frac{f(X_1) + \cdots + f(X_N)}{N} - \mathbb{E}(f(X_1))\right| \geq r\right) \leq 2\exp\left(-\frac{Nr^2}{2\|f\|^2_{\mathrm{Lip}}}\right).$$

The right hand side does not depend on n, in other words is uniform over the dimension. If $f(x) = \langle x, \theta\rangle$, $|\theta| = 1$, then $\sqrt{N}\left(\frac{f(X_1)+\cdots+f(X_N)}{N} - \mathbb{E}(f(X_1))\right) \sim \gamma_n$ and $\|f\|_{\mathrm{Lip}} = 1$.

Proof. The function $x \in (\mathbb{R}^n)^N = \mathbb{R}^{nN} \mapsto F(x) = \frac{1}{N}(f(x_1) + \cdots + f(x_N))$ is Lipschitz with

$$\|F\|_{\mathrm{Lip}} \leq \frac{\|f\|_{\mathrm{Lip}}}{N} \sup_{x \neq y} \frac{\sum_{i=1}^N |x_i - y_i|}{\sqrt{\sum_{i=1}^N |x_i - y_i|^2}} \leq \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}.$$

Moreover $\mathbb{E}(F(X_1, \ldots, X_N)) = \mathbb{E}(f(X_1))$ and $(X_1, \ldots, X_N) \sim \gamma_n^{\otimes N} = \gamma_{nN}$. $\square$

Remark (Logarithmic Sobolev inequalities)**.** *An examination of the proofs above reveals that the sub-Gaussian concentration inequalities are still valid when the Gaussian measure γ_n is replaced by any probability measure on $\mathbb{R}^n$ which satisfies a logarithmic Sobolev inequality. An advantage of using a logarithmic Sobolev inequality to deduce concentration of measure is that it is stable by tensor products, whereas concentration is not stable by tensor products. More precisely, and seen is Section 1.6, if μ satisfies the logarithmic Sobolev inequality on $\mathbb{R}^n$ then for any $n \geq 1$ the tensor product $\mu^{\otimes N}$ satisfies to exactly the same logarithmic Sobolev inequality stated on $\mathbb{R}^{nN}$. In other words the constant in front of the right hand side is dimension free. It turns out that it can be shown that dimension free sub-Gaussian concentration of measure for Lipschitz functions is in fact equivalent to a* W_2 *transpotation inequality. We will study transportation inequalities later on.*

Remark (Poincaré inequality and the exponential tail)**.** *It can be shown that the double sided exponential distribution on $\mathbb{R}$ with density $x \mapsto \frac{1}{2}\exp(-|x|)$ satisfies a Poincaré inequality. However it cannot satisfy a logarithmic Sobolev inequality due to its sub-exponentail tail which is not sub-Gaussian. The Herbst method allows to show that if a probability measure satisfies to a Poincaré inequality then this implies sub-exponential concentration around the mean for Lipschitz functions.*

3.2 Transportation inequalities

For any $p \geq 1$, let $\mathcal{P}_p(\mathbb{R}^n)$ be the set of probability measures on $\mathbb{R}^n$ with finite moment of order p in other words the set of probability measures μ on $\mathbb{R}^n$ such that $|\cdot|^p \in \mathbb{L}^1(\mu)$. The Wasserstein–Kantorovich distance on $\mathcal{P}_p(\mathbb{R}^n)$ is defined for any $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$ by

$$\mathrm{W}_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^p}{p}\, d\pi(x,y)\right)^{1/p},$$

where $\Pi(\mu, \mu)$ is the set of probability measures on the product space $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ and ν. The quantity $\mathrm{W}_p(\mu, \nu)$ is a "transportation cost" between μ and ν. It can be shown that W_p is a distance on $\mathcal{P}_p(\mathbb{R}^n)$, and that for any $(\mu_k)_{k\in\mathbb{N}}$ and μ in $\mathcal{P}_p(\mathbb{R}^n)$, we have $\mathrm{W}_p(\mu_k, \mu) \to 0$ if and only if $\mu_k \to \mu$ with respect to continuous test functions $f : \mathbb{R}^n \to \mathbb{R}$ such that $x \mapsto f(x)/(1 + |x|^p)$ is bounded.

The definition of W_2 above differs from the one used in Section 1.5 by a factor of $\sqrt{2}$.

Note that by Jensen's inequality, $\mathrm{W}_1 \leq p^{1/p}\mathrm{W}_p$.

The variational definition of W_p above involves a linear expression with respect to π optimized over an (infinite dimensional) convex constraint on π. Such a structure suggests duality. The Kantorovich–Rubinstein duality states that for any $p \geq 1$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$,

$$\mathrm{W}_p(\mu, \nu)^p = \sup \left(\int f \, \mathrm{d}\mu - \int g \, \mathrm{d}\nu \right)$$

where the supremum is taken over the set of bounded Lipschitz functions $f, g : \mathbb{R}^n \to \mathbb{R}$ such that $f(x) - g(y) \leq |x - y|^p/p$ for any $x, y \in \mathbb{R}^n$. In other words,

$$\mathrm{W}_p(\mu, \nu)^p = \sup \left(\int Q(f) \, \mathrm{d}\mu - \int f \, \mathrm{d}\nu \right)$$

where the supremum is taken over the set of bounded Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}$, and where $Q(f)$ is the "infimum convolution" of f with $\frac{1}{p} |\cdot|^p$ defined by

$$Q(f)(x) := \inf_{y\in\mathbb{R}^n} \left(f(y) + \frac{|x - y|^p}{p} \right), \quad x \in \mathbb{R}^n.$$

In the case $p = 1$, we have the simplified expression

$$\mathrm{W}_1(\mu, \nu) = \sup \left(\int f \, \mathrm{d}\mu - \int f \, \mathrm{d}\nu \right)$$

where the supremum is taken over the set of Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ with $\|f\|_{\mathrm{Lip}} \leq 1$.

For any fixed $p \geq 1$, we say that a probability measure $\mu \in \mathcal{P}_p(\mathbb{R}^n)$ satisfies a W_p transportation inequality if for some constant $C > 0$ and for any $\nu \in \mathcal{P}_p(\mathbb{R}^n)$,

$$W_p(\mu, \nu) \leq \sqrt{C\mathrm{H}(\nu \mid \mu)}.$$

The most used cases are $p \in \{1, 2\}$. This reminds the Pinsker–Csiszár–Kullback inequality

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sqrt{\frac{\mathrm{H}(\nu \mid \mu)}{2}}$$

where $\|\mu - \nu\|_{\mathrm{TV}} = \sup_A |\mu(A) - \nu(A)|$, which follows from the inequality $3(u - 1)^2 \leq (2u + 4)(u \log(u) - u + 1)$, $u \geq 0$. The total variation distance is a W_1 distance with respect to the atomic underlying distance $\mathbf{1}_{x\neq y}$ instead of the Euclidean distance $|x - y|$.

Katalin Marton has shown that the W_1 transportation inequality implies a sub-Gaussian concentration of measure for Lipschitz functions. The argument relies on the coupling expression of W_1. More precisely, let $A, B \subset \mathbb{R}^n$ be a couple of Borel subsets of $\mathbb{R}^n$ such that $\mu(A) > 0$ and $\mu(B) > 0$, and let $\mu_A = \mu(\cdot|A) = \mu(\cdot \cap A)/\mu(A)$ and $\mu_B = \mu_B(\cdot|B) = \mu(\cdot \cap B)/\mu(B)$ be the conditional distributions. The triangle inequality for W_1 gives

$$\mathrm{W}_1(\mu_A, \mu_B) \leq \mathrm{W}_1(\mu_A, \mu) + \mathrm{W}_1(\mu_B, \mu).$$

Now the W_1 transportation inequality $\mathrm{W}_1(\mu,\nu) \leq \sqrt{2c\mathrm{H}(\nu \mid \mu)}$ valid for any probability measure ν gives, using the notation $f_A = \mathbf{1}_A/\mu(A)$ and $f_B = \mathbf{1}_B/\mu(B)$,

$$\mathrm{W}_1(\mu_A,\mu_B) \leq \sqrt{2c\mathrm{Ent}_\mu(f_A)} + \sqrt{2c\mathrm{Ent}_\mu(f_B)} = \sqrt{-2c\log\mu(A)} + \sqrt{-2c\log\mu(B)}.$$

On the other hand, let us take now $B = (A_r)^c = \{x \in \mathbb{R}^n : \mathrm{dist}(x,A) \geq r\}$ for some $r \geq 0$. For such a couple A and B, the distance between the supports of the measures μ_A and μ_B is larger than or equal to r. Therefore, by the coupling variational formula for W_1,

$$\mathrm{W}_1(\mu_A,\mu_B) \geq r,$$

and thus if $r \geq r_* := \sqrt{-2c\log\mu(A)}$ then we get

$$\mu((A_r)^c) = \mu(B) \leq \exp\left(-\frac{(r-r_*)^2}{2c}\right).$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, then by the triangle inequality, for any $r > 0$,

$$\{f \leq \mu(f)\}_r \subset \{f < \mu(f) + r\|f\|_{\mathrm{Lip}}\}.$$

Taking $A = \{f \leq \mu(f)\}$ we obtain, for $r \geq r_*$, the sub-Gaussian bound

$$\mu(f-\mu(f) > r\|f\|_{\mathrm{Lip}}) \leq \exp\left(-\frac{(r-r_*)^2}{2c}\right).$$

The following theorem is due to Sergey Bobkov and Friedrich Götze. It states that the W_1 transportation inequality is the dual reformulation of the sub-Gaussian concentration for Lipschitz functions. The proof relies on the Kantorovich–Rubinstein dual or variational representation of W_1 as well as on a dual or variational representation of the entropy.

Theorem 3.4 (Transportation inequality W_1)**.** *For any $\mu \in \mathcal{P}_1(\mathbb{R}^n)$ and any constant $c > 0$, the following statements are equivalent:*

1. *Sub-Gaussian upper bound on Laplace transform of Lipschitz functions: for any Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ and any $\theta \in \mathbb{R}$,*

$$L(\theta) := \int \exp(\theta f)\,\mathrm{d}\mu \leq \exp\left(\frac{c}{2}\theta^2\|f\|_{\mathrm{Lip}}^2 + \theta\int f\mathrm{d}\mu\right);$$

2. *Transportation inequality W_1 for μ: for any $\nu \in \mathcal{P}_1(\mathbb{R}^n)$,*

$$\mathrm{W}_1(\nu,\mu) \leq \sqrt{2c\mathrm{H}(\nu \mid \mu)}.$$

In particular, this holds true when $\mu = \gamma_n$ with $c = 1$, for any $n \geq 1$.

Proof. We will use the following variational formula for the entropy, valid for any measurable $h : \mathbb{R}^n \to \mathbb{R}$ with $h \geq 0$, achieved for $g = \log(h/\mu(h))$ where $\mu(h) = \int h\mathrm{d}\mu$:

$$\mathrm{Ent}_\mu(h) = \sup\left\{\int hg\,\mathrm{d}\mu : \int \mathrm{e}^g\,\mathrm{d}\mu \leq 1\right\}.$$

Note that in 1. we can reduce to $\theta > 0$ by replacing f by $-f$ and also to $\mu(f) = 0$ and $\|f\|_{\mathrm{Lip}} = 1$ by translation and scaling. Now 1. for any $f : \mathbb{R}^n \to \mathbb{R}$ with $\|f\|_{\mathrm{Lip}} = 1$ rewrites

$$\int \mathrm{e}^g\,\mathrm{d}\mu \leq 1$$

with $g = \theta f - (c/2)\theta^2$, and the variational formula for the entropy yields

$$\int \left(\theta f - \frac{c}{2}\theta^2\right) h \, \mathrm{d}\mu \leq \mathrm{Ent}_\mu(h).$$

Conversely, if this inequality holds then we can check that we recover 1. by taking $h = \mathrm{e}^{\theta f - (c/2)\theta^2}$ since 1. follows from

$$\left(\int \mathrm{e}^{\theta f - (c/2)\theta^2} \, \mathrm{d}\mu\right) \log \int \mathrm{e}^{\theta f - (c/2)\theta^2} \, \mathrm{d}\mu \leq 0.$$

Now, by homogeneity, we can assume that h is a probability density function with respect to μ, and since f has zero mean for μ, we reformulate as

$$\int (fh - f) \, \mathrm{d}\mu \leq \frac{c}{2}\theta + \frac{1}{\theta} \int h \log h \, \mathrm{d}\mu.$$

Next, by taking the infimum over $\theta > 0$ we get

$$\int (fh - f) \, \mathrm{d}\mu \leq \sqrt{2c \int h \log h \, \mathrm{d}\mu}.$$

Denoting $\nu \ll \mu$ the probability measure such that $\mathrm{d}\nu/\mathrm{d}\mu = h$ and taking the supremum over f gives, thanks to the Kantorovich–Rubinstein dual formulation,

$$\mathrm{W}_1(\mu, \nu) \leq \sqrt{2c\mathrm{H}(\nu \mid \mu)},$$

which is 2. It remains to note that the argument can be reversed. $\square$

Remark (Variational formula for the entropy)**.** *The variational formula for the entropy comes from the fact that as a convex functional $f \geq 0 \mapsto \mathrm{Ent}_\mu(f)$, it can be represented as the envelope of its affine tangents, namely*

$$\begin{aligned}
\mathrm{Ent}_\mu(f) &= \sup_{g \geq 0} \Big\{ \int \Big(\log(g) - \log\Big(\int g \, \mathrm{d}\mu\Big)\Big) (f - g) \, \mathrm{d}\mu + \mathrm{Ent}_\mu(g) \Big\} \\
&= \sup_{g \geq 0} \Big\{ \int \Big(\log(g) - \log\Big(\int g \, \mathrm{d}\mu\Big)\Big) f \, \mathrm{d}\mu \Big\} \\
&= \sup_{\substack{g \geq 0 \\ \int g \, \mathrm{d}\mu = 1}} \Big\{ \int f \log(g) \, \mathrm{d}\mu \Big\} \\
&= \sup_{\int \mathrm{e}^g \, \mathrm{d}\mu = 1} \Big\{ \int fg \, \mathrm{d}\mu \Big\}.
\end{aligned}$$

Replacing g such that $\int \mathrm{e}^g \, \mathrm{d}\mu = 1$ by $g - \log \int \mathrm{e}^g \, \mathrm{d}\mu$ for an arbitry g we see that Ent_μ *is the Legendre transform of the log-Laplace transform in the sense that*

$$\mathrm{Ent}_\mu(f) = \sup_g \Big\{ \int fg \, \mathrm{d}\mu - \log \int \mathrm{e}^g \, \mathrm{d}\mu \Big\}.$$

And the entropy is the Legendre transform (convex dual) of the log-*Laplace transform*

$$\sup_{\substack{g \geq 0 \\ \int g \, \mathrm{d}\mu = 1}} \{ \int fg \, \mathrm{d}\mu - \mathrm{Ent}_\mu(g) \} = \log \int \mathrm{e}^f \, \mathrm{d}\mu.$$

The following theorem, due to Sergey Bobkov and Friedrich Götze, provides the dual reformulation of the W_2 transportation inequality, via infimum convolution.

Theorem 3.5 (Transportation inequality W_2). *For any $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ and any constant $c > 0$, the following statements are equivalent:*

1. *Transportation inequality* W_2 *for μ: for any $\nu \in \mathcal{P}_2(\mathbb{R}^n)$,*

$$\mathrm{W}_2(\mu,\nu) \le \sqrt{c\mathrm{H}(\nu \mid \mu)};$$

2. *For any bounded and Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\int \exp(Q_c(f))\,\mathrm{d}\mu \le \exp \int f\,\mathrm{d}\mu$$

where $Q_c(f)(x) := \inf_{y\in\mathbb{R}^n}(f(y) + \frac{|x-y|^2}{2c}) = c^{-1}Q_1(cf)$.

Note that $\mathrm{W}_1(\nu \mid \mu) \le \sqrt{2}\mathrm{W}_2(\nu \mid \mu)$ and therefore the W_2 transportation inequality implies the W_1 transportation inequality.

Note that the infimum convolution inequality 2. implies sub-Gaussian concentration for Lipschitz functions, namely that for any lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, and any $\theta \in \mathbb{R}$,

$$\int \exp(\theta f)\,\mathrm{d}\mu \le \exp\left(\frac{c}{2}\theta^2\|f\|_{\mathrm{Lip}}^2 + \int f\mathrm{d}\mu\right).$$

Indeed, after assuming without loss of generality as usual that $\mu(f) = 0$, $\theta = 1$, and that f is bounded, this follows from 2. together with the fact that for any $x \in \mathbb{R}^n$,

$$Q_c(f)(x) \ge f(x) + \inf_{y\in\mathbb{R}^n}\left(-\|f\|_{\mathrm{Lip}}|x-y| + \frac{|x-y|^2}{2c}\right) \ge f(x) - \frac{c}{2}\|f\|_{\mathrm{Lip}}^2.$$

Proof. Let us prove that 2.⇒1. For any bounded and Lipschitz g, 2. for $f = g/c$ gives

$$\int \exp\left(\frac{1}{c}\left(Q_1(g) - \int g\,\mathrm{d}\mu\right)\right)\mathrm{d}\mu \le 1.$$

Now, for any probability measure ν with density $f = \mathrm{d}\nu/\mathrm{d}\mu$ with respect to μ, the variational formula for the entropy $\mathrm{Ent}_\mu(f) = \sup\{\mu(fh) : \mu(\mathrm{e}^h) \le 1\}$ gives, for the special choice $h = \frac{1}{c}\left(Q_1(g) - \int g\,\mathrm{d}\mu\right)$,

$$\int \frac{1}{c}\left(Q_1(g) - \int g\,\mathrm{d}\mu\right) f\,\mathrm{d}\mu = \int hf\,\mathrm{d}\mu \le \mathrm{Ent}_\mu(f).$$

Taking the supremum over g gives 1. by the Kantorovich–Rubinstein duality.

Let us show that 1.⇒2. For any probability density function f with respect to μ, we have, denoting ν the probability measure such that $\mathrm{d}\nu = f\mathrm{d}\mu$,

$$\mathrm{W}_2(\mu,\nu)^2 \le c\mathrm{Ent}_\mu(f).$$

Using the Kantorovich–Rubinstein duality, for any bounded Lipschitz $g : \mathbb{R}^n \to \mathbb{R}$,

$$\int\left(Q_1(g) - \int g\,\mathrm{d}\mu\right) f\,\mathrm{d}\mu = \int Q_1(g) f\,\mathrm{d}\mu - \int g\,\mathrm{d}\mu \le c\int f\log f\,\mathrm{d}\mu.$$

Since g and $Q_1(g)$ are bounded, one can take

$$f \propto \exp\left(\frac{1}{c}\left(Q_1(g) - \int g \,\mathrm{d}\mu\right)\right)$$

and get

$$\log \int \exp\left(\frac{1}{c}Q_1(g) - \exp\left(\frac{1}{c}\int g \,\mathrm{d}\mu\right)\right) \leq 0,$$

which can be rewritten as 2. using $c^{-1}Q_1(g) = Q_c(c^{-1}g)$. □

Theorem 3.6 (W_2 inequality for Gauss measures)**.** *For any $n \geq 1$ and $\nu \in \mathcal{P}_2(\mathbb{R}^n)$,*

$$\mathrm{W}_2(\nu, \gamma_n) \leq \sqrt{\mathrm{H}(\nu \mid \gamma_n)}.$$

The first proof, due to Michel Talagrand, is by induction on n. We give here a proof due to Sergey Bobkov, Ivan Gentil, and Michel Ledoux, which is based on the logarithmic Sobolev inequality and a Hamilton-Jacobi equation for infimum convolutions.

Proof. Set $\mu = \gamma_n$ and $c = 1$. Thanks to the characterization of the W_2 transportation inequality (theorem 3.5), it is enough to show that for any bounded Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int \exp(Q_c(f)) \,\mathrm{d}\mu \leq \exp \int f \,\mathrm{d}\mu.$$

Now the idea is to try to exploit the logarithmic Sobolev inequality for μ in order to obtain a differential inequality for the Laplace transform of $Q_c(f)$, just like in the Herbst argument that we have already used to get from the logarithmic Sobolev inequality a sub-Gaussian upper bound for Laplace transforms of Lipschitz functions. This requires to control the gradient of $Q_c(f)$. It turns out that we can benefit from the following fact: if $f : \mathbb{R}^n \to \mathbb{R}$ is bounded and Lipschitz and if $\theta \in [0, 1]$ then the "infimum-convolution"

$$u(t, x) = Q_t(f)(x) = \inf_{y \in \mathbb{R}^n} \left(f(y) + \frac{|x - y|^2}{2t}\right), \quad t > 0,\ x \in \mathbb{R}^n$$

is a "Hopf-Lax" solution of the Hamilton-Jacobi equation[3]

$$\begin{cases} \partial_t u + \frac{1}{2}|\nabla_x u|^2 = 0 & \text{on } (0, \infty) \times \mathbb{R}^n, \\ u(0, \cdot) = f. \end{cases}$$

It follows that $g(x, \theta) := Q_c(\theta f)(x) = \theta Q_{c\theta}(f)(x)$ satisfies to

$$\theta \partial_\theta g = g - c\frac{1}{2}|\nabla_x g|^2.$$

Using this formula, we get from the logarithmic Sobolev inequality for μ

$$\mathrm{Ent}_\mu(h^2) \leq 2c \int |\nabla h|^2 \,\mathrm{d}\mu$$

with $h = \exp(\frac{1}{2}g)$ that

$$\theta L'(\theta) \leq L(\theta) \log L(\theta) \quad \text{where} \quad L(\theta) := \int \mathrm{e}^g \,\mathrm{d}\mu,$$

[3]See for instance Section 3.3 in the book "Partial Differential Equations" by Lawrence C. Evans.

in other words $K' \leq 0$ on $(0,1]$ where $K(\theta) := (1/\theta)\log L(\theta)$ for any $\theta \in [0,1]$. Now K is continuous on $[0,1]$ and thus $K(1) \leq K(0)$. Moreover, the identities $L(0) = 1$ and $K(0) = (\log L)'(0) = L'(0)/L(0) = \int f\,\mathrm{d}\mu$, which gives

$$\int \exp(Q_c(f))\,\mathrm{d}\mu = L(1) = \exp(K(1)) \leq \exp(K(0)) = \exp\int f\,\mathrm{d}\mu.$$

$\square$

The proof above is not Gaussian. It shows that if a probability measure μ on $\mathbb{R}^n$ satisfies a logarithmic Sobolev inequality then it satisfies also a W_2 transportation inequality.

It was proved few years ago by Nathael Gozlan probabilistically and then by Michel Ledoux analytically that the W_2 transportation inequality is actually equivalent to a dimension free sub-Gaussian concentration inequality for Lipschitz functions.

$$\begin{array}{rcccrc}
\displaystyle\int \mathrm{e}^{Q_c(f)}\mathrm{d}\mu \leq & \mathrm{e}^{\int f\mathrm{d}\mu} & \leftrightarrow & & \mathrm{W}_2(\nu,\mu) \leq & \sqrt{cH(\nu \mid \mu)} \\
\downarrow & & & & \downarrow & \\
\displaystyle\int \mathrm{e}^{f}\mathrm{d}\mu \leq & \mathrm{e}^{\frac{c}{2}\|f\|_{\mathrm{Lip}}^2+\int f\mathrm{d}\mu} & \leftrightarrow & & \mathrm{W}_1(\nu,\mu) \leq & \sqrt{2cH(\nu \mid \mu)} \\
\nwarrow & & & & \nearrow & \\
 & & \multicolumn{2}{c}{\displaystyle\mathrm{Ent}_\mu(f) \leq \frac{c}{2}\int \frac{|\nabla f|^2}{f}\,\mathrm{d}\mu} & &
\end{array}$$

Chapter 4

Isoperimetric inequalities

The classical isoperimetric inequality for the Lebesgue measure on $\mathbb{R}^n$ states that among the Borel sets of given volume, the boundary measure is minimal for balls. Similarly, the classical isoperimetric inequality for the uniform distribution on the sphere states that among the Borel sets on the sphere of given area, the boundary measure is minimal for sphere caps. This section is devoted to the isoperimetry for the Gaussian measure γ_n, for which it turns out that the extremal sets are half-spaces. This can be seen as a consequence of the isoperimetric inequality on the sphere thanks to the Poincaré observation.

For a Borel measure μ on $\mathbb{R}^n$, the boundary measure of a Borel set $A \subset \mathbb{R}^n$ is

$$\mu^+(A) = \liminf_{r \to 0} \frac{\mu(A_r) - \mu(A)}{r}$$

where $A_r := \{x \in \mathbb{R} : \mathrm{dist}(x, A) < r\}$ is the r-neightborhood of $A \subset \mathbb{R}^n$.

The probability density function and the cumulative distribution function of the standard Gauss measure γ_1 on $\mathbb{R}$ are given, for any $x \in \mathbb{R}$, by

$$\varphi(x) = (2\pi)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}x^2} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t.$$

Note that for $x \geq 0$, $\Phi(x) = \frac{1+\mathrm{erf}(\sqrt{2}x)}{2}$ where $\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\mathrm{d}t$.

Lemma 4.1 (Halfspaces)**.** *If $H = (-\infty, a]$, $a \in \mathbb{R}$, then*

$$\gamma_1(H) = \Phi(a) \quad \textit{and} \quad \gamma_1^+(H) = \varphi(a).$$

More generally, for any $n \geq 1$, $u \in \mathbb{R}^n$ with $|u| = 1$, and $a \in \mathbb{R}$, the (affine) halfspace

$$H := \{x \in \mathbb{R}^n : \langle u, x \rangle \leq a\}$$

satisfies

$$\gamma_n(H) = \Phi(a) \quad \textit{and} \quad \gamma_n^+(H) = \varphi(a).$$

In particular, for all $p \in (0,1)$, a halfspace H of measure $\gamma_n(H) = p$ has boundary measure

$$\gamma_n^+(H) = \varphi \circ \Phi^{-1}(p) =: I(p).$$

This dimension free phenomenon comes from the rotational invariance.

Proof. The rotational invariance of γ_n reduces the problem to $n = 1$, and in this case $\gamma_1(H) = \Phi(a)$ while the fact that $\varphi = \Phi'$ gives

$$\gamma_1(H_r) - \gamma_1(H) = \Phi(a + r) - \Phi(a) = r\varphi(a) + o(r)$$

which provides $\gamma_1^+(H) = \varphi(a)$. □

Let us define $I : [0,1] \mapsto [0,\infty)$ by $I(0) = I(1) = 0$ and for any $p \in (0,1)$,

$$I(p) := (\varphi \circ \Phi^{-1})(p).$$

Lemma 4.2 (Gaussian isoperimetric function). *The function $I : [0,1] \to [0,\infty)$ is concave, symmetric with respect to $1/2$ where it achieves its maximum $I(1/2) = (2\pi)^{-1/2}$ and satisfies to the differential equation $II'' = -1$ on $(0,1)$. Furthermore*

$$I(p) \underset{p\to 0}{\sim} p\sqrt{-2\log p} \quad \text{and} \quad \lim_{p\to 0} I'(p) = +\infty.$$

Proof. On $(0,1)$, $\Phi' = \varphi$ gives $(\Phi^{-1})' = 1/(\varphi \circ \Phi^{-1}) = 1/I$ and $\varphi'(x) = -x\varphi(x)$ gives

$$I' = \frac{\varphi' \circ \Phi^{-1}}{I} = -\Phi^{-1} \quad \text{and} \quad I'' = -\frac{1}{\varphi \circ \Phi^{-1}} = -\frac{1}{I}.$$

This gives $II'' = -1$, the concavity of I, and $\lim_{p\to 0} I'(p) = \lim_{p\to 0} -\Phi^{-1}(p) = +\infty$. Finally, since φ is even, it follows that $\Phi - \Phi(0) = \Phi - 1/2$ is odd, and therefore I is symmetric with respect to $1/2$. Since I is concave, its maximum is $I(1/2) = \varphi(0) = (2\pi)^{-1/2}$. Finally $\Phi(x) \underset{x\to-\infty}{\sim} -\frac{\varphi(x)}{x}$ gives $\Phi^{-1}(p) \underset{p\to 0}{\sim} -\sqrt{-2\log(p)}$ and then $I(p) \underset{p\to 0}{\sim} p\sqrt{-2\log p}$. □

Theorem 4.3 (Gaussian isoperimetry). *For any $p \in [0,1]$, among all Borel subsets $A \subset \mathbb{R}^n$ such that $\gamma_n(A) = p$, the boundary measure $\gamma_n^+(A)$ is minimal for halfspaces:*

$$I(p) = \inf_{\substack{A\subset\mathbb{R}^n \\ \gamma_n(A)=p}} \gamma_n^+(A),$$

in other words for any Borel set $A \subset \mathbb{R}^n$,

$$I(\gamma_n(A)) \leq \gamma_n^+(A); \tag{4.1}$$

In other words, for any Borel subset $A \subset \mathbb{R}^n$ and any halfspace H, we have

$$\gamma_n(A) = \gamma_n(H) \quad \Rightarrow \quad \gamma_n^+(H) \leq \gamma_n^+(A).$$

It possible to deduce (4.1) from the isoperimetric inequality for the uniform distribution on spheres by letting the dimension and radius tend to infinity, as explained by Michel Ledoux in [23]. Antoine Ehrhart gave in [17] another direct proof of (4.1) which relies on symmetrization. Christer Borell gave an alternative proof in [8].

FIXME: explain why half-spaces are in some sense the unique extremal sets.

Theorem 4.4 (Gaussian isop. – Integrated version). *For all Borel set $A \subset \mathbb{R}^n$ and $r > 0$,*

$$\gamma_n(A_r) \geq \Phi(\Phi^{-1}(\gamma_n(A)) + r). \tag{4.2}$$

It is possible to deduce (4.1) from (4.2) immediately since

$$\gamma_n^+(A) = \liminf_{r\to 0} \frac{\gamma_n(A_r) - \gamma_n(A)}{r} \geq \Phi'(\Phi^{-1}(\gamma_n(A))) = I(\gamma_n(A)).$$

Conversely, to deduce (4.2) from (4.1), we may assume first by approximation that A is a finite union of open balls. The familly of such sets is closed under the r-neighborhood operation, and thus the $\underline{\lim}$ in the definition of γ_n^+ is a true limit. In this case $\gamma_n^+(A)$ is the integral of the density of γ_n along the boundary ∂A which is a union of spheres, in other words $\partial_r \gamma_n(A_r) = \gamma_n^+(A_r)$. Now

$$a(r) := \Phi^{-1}(\gamma_n(A_r)) = a(0) + \int_0^r a'(t)\,\mathrm{d}t,$$

and $a'(t) = \gamma_n^+(A_t)/(\varphi \circ \Phi^{-1})(A_t) \geq 1$ by (4.1), thus $a(r) \geq a(0) + r$, hence (4.2).

Remark. *If $\gamma_n(A) = p \geq 1/2$ and $a = \Phi^{-1}(p)$ then $a > 0$ and* (4.2) *gives, for all $r > 0$,*

$$\gamma_n(A_r) \geq \Phi(a+r) \geq \Phi(r) \geq 1 - \mathrm{e}^{-\frac{1}{2}r^2},$$

and this bound is remarkably dimension free!

Theorem 4.5 (Gaussian Bobkov isoperimetry)**.** *For every smooth function $f : \mathbb{R}^n \to [0,1]$,*

$$I\left(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma_n\right) - \int_{\mathbb{R}^n} I(f)\,\mathrm{d}\gamma_n \leq \int_{\mathbb{R}^n} |\nabla f|\,\mathrm{d}\gamma_n; \tag{4.3}$$

The inequality (4.3) was introduced and proved by Sergey Bobkov in [6]. It is the first known functional form of the Gaussian isoperimetric inequality. It is possible to deduce (4.1) from (4.3) by approximating indicators with smooth functions, see [22, 7], and conversely, one can deduce (4.3) form (4.1), see for instance [7]. The Bobkov inequality (4.3) plays for the Gaussian the role played by the optimal L^1 Sobolev inequality for the Lebesgue measure. We refer to [5] and [29] for a discussion on isoperimetric inequalities. Here « L^1 » means that the functional inequality involves the L^1 norm of the gradient.

Direct proof of (4.3) *using semigroup interpolation.* We follow the proof given by Dominique Bakry and Michel Ledoux in [2]. Let $(P_t)_{t\geq 0}$ be the Ornstein–Uhlenbeck semigroup. Let $\varepsilon \in (0,1)$ and $f : \mathbb{R}^n \to [\varepsilon, 1-\varepsilon]$ be smooth. We get, for all $t > 0$,

$$\begin{aligned}
I(P_t f) - P_t(I(f)) &= -\int_0^t \partial_s \, P_s(I(P_{t-s}f))\mathrm{d}s \\
&= -\int_0^t P_s(I''(P_{t-s}f)\,|\nabla P_{t-s}f|^2)\mathrm{d}s \\
&\leq -\int_0^t P_s(I''(P_{t-s}f)\,|\nabla P_{t-s}f|\,\mathrm{e}^{-(t-s)}P_{t-s}|\nabla f|)\mathrm{d}s \\
&\leq \int_0^t \mathrm{e}^{-(t-s)}\alpha(t-s)P_s(P_{t-s}|\nabla f|)\mathrm{d}s \\
&= \left(\int_0^t \mathrm{e}^{-s}\alpha(s)\mathrm{d}s\right) P_t(|\nabla f|)
\end{aligned}$$

where (recall that $II'' = -1$)

$$\alpha(u) := \sup_{\mathbb{R}^n}(-I''(P_u f)|\nabla P_u f|) = \sup_{\mathbb{R}^n}(I(P_u f)|\nabla P_u f|).$$

To perform the computation above, we have used a couple of now familiar ingredients:

- The diffusion formula $L(\beta(h)) = \beta'(h)Lh + \beta''(h)|\nabla h|^2$ giving, with $\beta = I$, $g := P_{t-s}f$,

$$\partial_s P_s(I(g)) = P_s(L(I(g))) - P_s(I'(g)Lg) = P_s(I''(g)|\nabla g|^2).$$

- The sub-commutation $|\nabla P_u f| \leq \mathrm{e}^{-u}P_u(|\nabla f|)$ with $u = t-s$.

It remains to find an upper bound for the α above. Using again the same semigroup interpolation, same familiar ingredients, and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
[I(P_t f)]^2 &\geq [I(P_t f)]^2 - [P_t(I(f))]^2 \\
&= -\int_0^t \partial_s [P_s(I(P_{t-s}f))]^2 \,\mathrm{d}s
\end{aligned}$$

$$
\begin{aligned}
&= -2\int_0^t P_s(I(P_{t-s}f))P_s\left(I''(P_{t-s}f)\,|\nabla P_{t-s}f|^2\right)\mathrm{d}s \\
&= 2\int_0^t P_s(I(P_{t-s}f))P_s\left(\frac{|\nabla P_{t-s}f|^2}{I(P_{t-s}f)}\right)\mathrm{d}s \\
&\geq 2\int_0^t [P_s(|\nabla P_{t-s}f|)]^2\,\mathrm{d}s \\
&\geq 2\int_0^t \mathrm{e}^{2s}|\nabla P_s(P_{t-s}f)|^2\mathrm{d}s \\
&\geq (\mathrm{e}^{2t}-1)|\nabla P_t f|^2,
\end{aligned}
$$

therefore

$$
\alpha(u) = \sup_{\mathbb{R}^n}\frac{|\nabla P_u f|}{I(P_u f)} \leq \frac{1}{\sqrt{\mathrm{e}^{2u}-1}}\frac{I(P_u f)}{I(P_u f)} = \frac{1}{\sqrt{\mathrm{e}^{2u}-1}},
$$

which gives finaly

$$
\begin{aligned}
I(P_t f) - P_t(I(f)) &= \left(\int_0^t \mathrm{e}^{-s}\alpha(s)\mathrm{d}s\right)P_t(|\nabla f|) \\
&\leq \left(\int_0^t \frac{\mathrm{e}^{-2s}}{\sqrt{1-\mathrm{e}^{-2s}}}\mathrm{d}s\right)P_t(|\nabla f|) \\
&= \left(\int_0^t (\sqrt{1-\mathrm{e}^{-2s}})'\mathrm{d}s\right)P_t(|\nabla f|) \\
&= \sqrt{1-\mathrm{e}^{-2t}}P_t(|\nabla f|).
\end{aligned}
$$

Of course $P_t(\cdot)(x) = \mathcal{N}(x\mathrm{e}^{-t}, 1-\mathrm{e}^{-2t}) \to \gamma_n$ as $t\to\infty$, for any $x\in\mathbb{R}^n$. □

Theorem 4.6 (Gaussian Bobkov isoperimetry bis)**.** *For all smooth $f:\mathbb{R}^n\to[0,1]$,*

$$
I\left(\int_{\mathbb{R}^n} f\,\mathrm{d}\gamma_n\right) \leq \int_{\mathbb{R}^n}\sqrt{I^2(f)+|\nabla f|^2}\,\mathrm{d}\gamma_n. \tag{4.4}
$$

As for (4.3), it is possible to deduce (4.1) from (4.4) by approximating indicators with smooth functions, see [22, 7], and conversely, one can deduce (4.4) form (4.1), see [7].

It is immediate to deduce (4.3) from (4.4) using $\sqrt{a^2+b^2}\leq|a|+|b|$. The advantage of (4.4) over (4.3) is that (4.4) can be tensorized. The inequality (4.4) was invented and proved by Sergey Bobkov [7]. In the sequel, we give three proofs of this inequality.

It is important to understand that the logarithmic Sobolev inequality for γ_n with its optimal constant can be deduced from (4.4), namely by taking $f=\varepsilon g^2$ with $\int_{\mathbb{R}^n} g^2\,\mathrm{d}\gamma_n = 1$ in (4.4) and using $I(p)\sim_{p\to0} p\sqrt{-2\log p}$, we get

$$
\int_{\mathbb{R}^n} g^2\log(g^2)\,\mathrm{d}\gamma_n \leq 2\int_{\mathbb{R}^n}|\nabla g|^2\,\mathrm{d}\gamma_n.
$$

Indeed, denoting for short $\mu(h):=\int h\mathrm{d}\gamma_n$, we have $\mathrm{I}(\mu(f)) \underset{\varepsilon\to0}{\sim} \varepsilon\sqrt{-2\log(\varepsilon)}$ while

$$
\begin{aligned}
\int_{\mathbb{R}^n}\sqrt{\mathrm{I}(f)^2+|\nabla f|^2}\,\mathrm{d}\gamma_n &\underset{\varepsilon\to0}{\sim} \int_{\mathbb{R}^n}\left(-\varepsilon^2 g^4 2\log(\varepsilon g^2)+\varepsilon^2 4|\nabla g|^2 g^2\right)\mathrm{d}\gamma_n \\
&= \varepsilon\sqrt{-2\log(\varepsilon)}\int_{\mathbb{R}^n}\left(g^2\sqrt{1-2\frac{\log(g^2)}{-2\log(\varepsilon)}+4\frac{|\nabla g|^2}{-2\log(\varepsilon)g^2}}\right)\mathrm{d}\gamma_n \\
&\underset{\varepsilon\to0}{\sim} \varepsilon\sqrt{-2\log(\varepsilon)}\int_{\mathbb{R}^n}\left(g^2\left(1-\frac{\log(g^2)}{-2\log(\varepsilon)}+2\frac{|\nabla g|^2}{-2\log(\varepsilon)g^2}\right)\right)\mathrm{d}\gamma_n
\end{aligned}
$$

$$= \varepsilon\sqrt{-2\log(\varepsilon)}\left(1 - \frac{1}{-2\log(\varepsilon)}\int_{\mathbb{R}^n} g^2 \log(g^2)\,\mathrm{d}\gamma_n + 2\frac{1}{-2\log(\varepsilon)}\int_{\mathbb{R}^n} |\nabla g|^2\,\mathrm{d}\gamma_n\right).$$

This argument, due to William Beckner, mirrors the linearization argument $g = 1 + \varepsilon f$ that allows to get the Poincaré inquality from the logarithmic Sobolev inequality. It is also important to mention that it possible to recover directly the sub-Gaussian concentration of measure for Lipschitz functions under γ_n from the sets formulation of the Gaussian isoperimetry (4.1), see [2, 4].

Direct proof of (4.4) *using semigroup interpolation.* We give a proof due to Dominique Bakry and Michel Ledoux (see [2] or the simpler [23]). Let $(P_t)_{t\geq 0}$ be the Ornstein–Uhlenbeck semigroup, and L be its infinitesimal generator. Let $\varepsilon \in (0,1)$ and $f : \mathbb{R}^n \to [\varepsilon, 1-\varepsilon]$ be smooth. Fix $t \geq 0$, and set, for any $t \geq 0$,

$$\alpha(t) := \int \sqrt{I^2(P_t f) + |\nabla P_t f|^2}\,\mathrm{d}\gamma_n.$$

It suffices to show that $\alpha(\infty) \leq \alpha(t)$, via $\alpha' \leq 0$. Now, setting $f_t := P_t f$,

$$\alpha'(s) = \int \frac{(II')(f_t)Lf_t + \langle \nabla f_t, \nabla L f_t\rangle}{\sqrt{I^2(f_t) + c|\nabla f_t|^2}}\,\mathrm{d}\gamma_n = (*) + (**).$$

Let us remove L in (*) and (**) by using an integration by parts. Namely, for (*), by integration by parts, using $II'' = -1$ and setting $K(f_t) := I^2(f_t) + |\nabla f_t|^2$,

$$\begin{aligned}(*) := \int \frac{(II')(f_t)Lf_t}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n &= -\int \langle \nabla f_t, \nabla \frac{(II')(f_t)}{\sqrt{K(f_t)}}\rangle\,\mathrm{d}\gamma_n \\ &= -\int \frac{(I'^2 - 1)(f_t)|\nabla f_t|^2}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n \\ &\quad + \int \frac{(II')^2(f_t)|\nabla f_t|^2 + (II')(f_t)\langle \nabla f_t, \nabla^2 f_t \nabla f_t\rangle}{(K(f_t))^{3/2}}\,\mathrm{d}\gamma_n.\end{aligned}$$

On the other hand, for (**), using $\langle \nabla f_t, \nabla L f_t\rangle = \langle \nabla f_t, L\nabla f_t\rangle - |\nabla f_t|^2$, we get

$$(**) := \int \frac{\langle \nabla f_t, \nabla L f_t\rangle}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n = \int \frac{\langle \nabla f_t, L\nabla f_t\rangle}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n - \int \frac{|\nabla f_t|^2}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n,$$

and the term with L can be rewritten, using an integration by part, as

$$\begin{aligned}\int \frac{\langle \nabla f_t, L\nabla f_t\rangle}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n &= -\int \frac{\|\nabla^2 f_t\|_2^2}{\sqrt{K(f_t)}}\,\mathrm{d}\gamma_n \\ &\quad + \int \frac{(II')(f_t)\langle \nabla f_t, \nabla^2 f_t \nabla f_t\rangle + \langle \nabla f_t, (\nabla^2 f_t)^2 \nabla f_t\rangle}{(K(f_t))^{3/2}}\,\mathrm{d}\gamma_n\end{aligned}$$

where $\|\nabla^2 f_t\|_2^2 := \sum_{j,k=1}^n (\partial_{jk}^2 f_t) = \mathrm{Trace}(\nabla_t^2 f \nabla^2 f_t)$ is the square of the Hilbert-Schmidt or trace or Frobenius norm of the Hessian matrix $\nabla^2 f_t$ of f_t.

Gathering all the terms gives, after some algebra,

$$\alpha'(t) = \int \frac{R(f_t)}{K^{3/2}(f_t)}\,\mathrm{d}\gamma_n$$

where, omitting f_t in the formula,

$$R = -I'^2|\nabla|^4 + 2II'\langle \nabla, \nabla^2\nabla\rangle - I^2\|\nabla^2\|_2^2 - \|\nabla^2\|_2^2|\nabla|^2 + \langle \nabla, (\nabla^2)^2\nabla\rangle$$

$$= -\mathrm{Trace}((I'\nabla\nabla^\top - I\nabla^2)(I'\nabla\nabla^\top - I\nabla^2)) + Q$$

where

$$Q := \langle \nabla, (\nabla^2)^2 \nabla \rangle - \|\nabla^2\|_2^2 |\nabla|^2 \leq 0.$$

□

Direct proof of (4.4) *using stochastic calculus.* We give the proof due to Franck Barthe and Bernard Maurey [4] inspired by a bit more abstract work of Mireille Capitaine, Elton Hsu, and Michel Ledoux [12]. By regularization, one can reduce to the case where f is smooth. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion of $\mathbb{R}^n$ and let $(\mathcal{F}_t)_{t\geq 0}$ be its natural filtration. For any $t \geq 0$, set

$$\begin{aligned} M_t &:= \mathbb{E}(f(B_1) \mid \mathcal{F}_t) = M_0 + \int_0^t m_s \, dB_s, \\ N_t &:= \mathbb{E}(\nabla f(B_1) \mid \mathcal{F}_t) = N_0 + \int_0^t n_s \, dB_s, \\ A_t &:= t \wedge 1 = A_0 + \int_0^t a_s \mathrm{d}s, \end{aligned}$$

(these processes are constant for $t \geq 1$). It suffices to show that the process

$$\left(\sqrt{I^2(M_t) + A_t |N_t|^2} \right)_{t \geq 0}$$

is a submartingale, since in that case

$$\begin{aligned} I(\mathbb{E}(f(B_1))) = \sqrt{I^2(\mathbb{E}(f(B_1)))} = \mathbb{E}F(0) & \\ &\leq \mathbb{E}F(1) = \mathbb{E}\sqrt{I^2(f(B_1)) + |\nabla f(B_1)|^2}. \end{aligned}$$

Let us assume for notational simplicity that $n = 1$. Let $J : \mathbb{R} \to \mathbb{R}$ be a continuous $\mathcal{C}^2$ function, contant outside $[0, 1]$, and such that $J(x) = I(x)$ when $x \in [\varepsilon, 1 - \varepsilon]$. Let $F(x, y, t) = \sqrt{J^2(x) + ty^2}$. Now

$$\partial_x F = \frac{2J(x)J'(x) + ty^2}{2F} \quad \text{and} \quad \partial_y F = \frac{J^2(x) + 2ty}{2F}$$

and thus

$$\begin{aligned} \partial^2_{x,x} F &= \frac{tJ'^2(x)y^2 + J^3(x)J''(x) + t|y|^2 J(x)J''(x)}{F^3}, \\ \partial^2_{x,y} F, &= -\frac{tyJ(x)J'(x)}{F^3}, \\ \partial^2_{y,y} F &= \frac{tJ^2(x)}{F^3}, \\ \partial_t F &= \frac{y^2}{2F} \end{aligned}$$

Now, Itô's formula gives, with $X_t := (M_t, N_t, A_t)$,

$$Y_t := F(X_t) = F(X_0) + \int_0^t (\partial_x F(X) \, dM + \partial_y F(X) \, dN) + \int_0^t \Delta(s) \mathrm{d}s$$

with

$$\Delta(s) := \partial_t F(X_s)a_s + \frac{1}{2}\left(\partial^2_{x,x}F(X_s)m_s^2 + 2\partial^2_{x,y}F(X_s)m_s n_s + \partial^2_{y,y}F(X_s)n_s^2\right).$$

The stochastic integral above has a bounded integrand, and is then a martingale. For the last integral above, we note that $(J(M_t))_{t\geq 0}$ and $(I(M_t))_{t\geq 0}$ coincide, which allows to use the relation $II'' = -1$, and this gives, omitting the variables,

$$2F^3\Delta = F^2N^2a + (I'^2AN^2 - I^2 - AN^2)m^2 - 2(II'AN)mn + I^2An^2.$$

Let us admit that $N^2a \geq m^2$. This gives $F^2N^2a \geq (I^2 + AN^2)m^2$, and then

$$2F^3\Delta \geq A(I'^2N^2m^2 - 2II'Nmn + In^2) = A(I'Nm - In)^2 \geq 0,$$

and therefore $(X_t)_{t\geq 0}$ is a submartingale.

It remains to show that $N^2a \geq m^2$. We have actually $a = 1$, and

$$M_t = \mathbb{E}(f(B_1 - B_t + B_t) \mid \mathcal{F}_t) = \alpha(t, B_t) \quad \text{and} \quad N_t = \partial_x\alpha(t, B_t)$$

where $\alpha(t,x) := \mathbb{E}(f(B_1 - B_t + x))$. But $(M_t)_{t\geq 0}$ is a martingale, and thus

$$dM_t = \partial_x\alpha(t, M_t)dB_t = N_t dB_t$$

by Itô's formula, which gives $m_s = N_s$, hence $m^2 = N^2$. □

4.1 Proof of the second Bobkov inequality

Inspired by the work [31] of Michel Talagrand, Sergey Bobkov gave in [7] a proof of the isoperimetric inequality (4.4), by using a tensorization of an inequality on the two-points space and the Central Limit Theorem. The starting point is the following inequality on the two-points space: for any $g : \{0,1\} \to [0,1]$,

$$I(\mathbb{E}_\mu(g)) \leq \mathbb{E}_\mu(\sqrt{I^2(g) + (\mathrm{D}g)^2})$$

where $(\mathrm{D}g)^2 = (g(1) - g(0))^2$. In other words, with $a := g(0)$ and $b := g(1)$,

$$I\left(\frac{a+b}{2}\right) \leq \frac{1}{2}\sqrt{I^2(a) + \left(\frac{b-a}{2}\right)^2} + \frac{1}{2}\sqrt{I^2(b) + \left(\frac{b-a}{2}\right)^2}.$$

FIXME:

Chapter 5

Bakry-Émery criterion

5.1 Gamma calculus

Let (X_t) be a Markov process, we let M be the state space, (P_t) be the semigroup and L be the generator. Given a function f and $t > 0$ we consider

$$\beta(s) = P_s((P_{t-s}f)^2), \quad s \in [0, t].$$

By Jensen's inequality β is increasing: $\beta(s + \varepsilon) \geq P_s((P_\varepsilon P_{t-s-\varepsilon}f)^2) = \beta(s)$. Observe that

$$\beta(t) - \beta(0) = P_t(f^2) - (P_t f)^2$$

evaluated at $x \in M$ is the variance of f with respect to $\delta_x P_t = P_t(\cdot)(x)$. Next, we have

$$\begin{aligned}
\beta'(s) &= \partial_s P_s(g^2) + P_s(2(\partial_s g)g) \\
&= P_s(L(g^2) - 2gLg) \\
&= 2P_s(\Gamma(g))
\end{aligned}$$

where $g = P_{t-s}f$ and

$$\Gamma(g) = \frac{1}{2}(L(g^2) - 2gLg).$$

The operator Γ is called *carré du champ*. More generally we introduce the bilinear form

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

Observe that

$$\Gamma(f, f) = \Gamma(f) \geq 0.$$

Indeed since β is increasing $\beta'(0) = 2\Gamma(P_t f) \geq 0$. Letting $t \to 0$ yields the positivity of Γ. Let us differentiate β once more. Recall that $g = P_{t-s}f$, since Γ is a bilinear form we have

$$\partial_s \Gamma(g) = 2\Gamma(\partial_s g, g) = -2\Gamma(Lg, g).$$

We thus obtain

$$\beta''(s) = 2\, P_s\left(L\Gamma(g) - 2\Gamma(Lg, g)\right).$$

It is thus tempting to introduce the iterated version of the carré du champ

$$\Gamma_2(f) = \frac{1}{2}\left(L\Gamma(f) - 2\Gamma(Lf, f)\right).$$

With this formalism we thus have

$$\beta''(s) = 4P_s(\Gamma_2(P_{t-s}f)).$$

For instance, if $X_t = \sqrt{2}B_t$, where (B_t) is a standard Brownian motion then $Lf = \Delta f$, from which we easily get (easily checked first in the unidimensional case $n = 1$)

$$\Gamma(f) = |\nabla f|^2 \quad \text{and} \quad \Gamma_2(f) = \frac{1}{2}\Delta|\nabla f|^2 - \langle \nabla f, \nabla \Delta f\rangle = \sum_{ij}(\partial_{ij}f)^2 = \mathrm{Tr}((\mathrm{Hess} f)^2).$$

Notice in particular that $\Gamma_2(f) \geq 0$. This already shows that $\beta'' \geq 0$ in that case. So β' is non decreasing, in particular $\beta'(0) \leq \beta'(t)$ In other words

$$|\nabla P_t f|^2 \leq P_t(|\nabla f|^2). \tag{5.1}$$

Similarly, if (X_t) is the Ornstein–Uhlenbeck semigroup then $Lf = \Delta f - \langle x, \nabla f\rangle$,

$$\Gamma(f) = |\nabla f|^2 \quad \text{and} \quad \Gamma_2(f) = \frac{1}{2}L|\nabla f|^2 - \langle \nabla f, \nabla L f\rangle = \sum_{ij}(\partial_{ij}f)^2 + |\nabla f|^2.$$

Notice that in that case we get $\Gamma_2(f) \geq \Gamma(f)$. This implies that

$$\beta''(s) \geq 2\beta'(s).$$

This differential inequality is easily integrated (Gronwall's lemma) and we get in particular

$$\beta'(t) \geq \mathrm{e}^{2t}\beta'(0).$$

In other words

$$|\nabla P_t f|^2 \leq \mathrm{e}^{-2t} P_t(|\nabla f|^2).$$

Of course we have seen before that this commutation property was an easy consequence of the Mehler formula. Similarly (5.1) is easily derived from the explicit expression of the heat semigroup. The point of the Γ calculus, is that it allows to derive such commutation properties for semigroups which do not have an explicit expression.

Definition 5.1 (Curvature – dimension inequality)**.** We say that the Markov process (X_t) satisfies the *curvature dimension inequality* $\mathrm{CD}(\rho, n)$ if for any function f we have

$$\Gamma_2(f) \geq \rho\Gamma(f) + \frac{1}{n}(Lf)^2.$$

Remark. *We have stated the curvature dimension condition in full generality but actually in this notes we shall only address the condition* $\mathrm{CD}(\rho, \infty)$*, for which the inequality reads*

$$\Gamma_2(f) \geq \rho\Gamma(f).$$

On the other hand $L = \Delta$ *satisfies to* $\mathrm{CD}(0, n)$*, in other words* $\mathrm{Tr}((\mathrm{Hess})^2) \geq \frac{1}{n}\Delta^2$*, since by the Cauchy – Schwarz inequality for* $\langle A, B\rangle = \mathrm{Tr}(AB)$ *on real symmetric matrices,*

$$\Delta = \mathrm{Tr}(\mathrm{Hess}) = \mathrm{Tr}(I_n \mathrm{Hess}) = \langle I_n, \mathrm{Hess}\rangle \leq \sqrt{\mathrm{Tr}(I_n)}\sqrt{\mathrm{Tr}(\mathrm{Hess}^2)} = \sqrt{n\mathrm{Tr}(\mathrm{Hess}^2)}.$$

Proposition 5.2 (Weak commutation)**.** *Let* $\rho \in \mathbb{R}$*, the following are equivalent:*

1. *The semi group* (P_t) *satisfies* $\mathrm{CD}(\rho, \infty)$*;*
2. *For every function* f *and every* t *we have* $\Gamma(P_t f) \leq \mathrm{e}^{-2\rho t} P_t\Gamma(f)$*.*

Proof. As seen above, letting $\alpha(s) = P_s(\Gamma(P_{t-s}f))$ we have $\alpha'(s) = 2P_s(\Gamma_2(P_{t-s}f))$. So under $\mathrm{CD}(\rho, \infty)$ we get $\alpha'(s) \geq 2\rho\alpha(s)$. Then by Gronwall lemma we get $\alpha(t)\mathrm{e}^{-2\rho t} \geq \alpha(0)$, which is the second inequality. Conversely, if the second assertion holds for all t then $\alpha(s) \geq \mathrm{e}^{2\rho s}\alpha(0)$ for every $s \in [0, t]$. A Taylor expansion of both expressions at $s = 0$ gives $\alpha'(0) \geq 2\rho\alpha(0)$, hence $\Gamma_2(P_t f) \geq \rho\Gamma(P_t f)$. Then letting $t \to 0$ we get $\mathrm{CD}(\rho, \infty)$. □

5.2 Poincaré inequality

In this section, we assume that there is a stationary distribution μ and that it is ergodic, in the sense that

$$P_t f \to \int_M f \, \mathrm{d}\mu,$$

for every f. The *Dirichlet form* is the quadratic form given by

$$\mathcal{E}(f,g) = \int_M \Gamma(f,g) \, \mathrm{d}\mu,$$

for all f, g. Again when $f = g$ we write $\mathcal{E}(f,f) = \mathcal{E}(f)$. Notice that by stationarity we have $\int_M L(f^2) \, \mathrm{d}\mu = 0$ and thus

$$\mathcal{E}(f) = \int_M \Gamma(f) \, \mathrm{d}\mu = -\int_M f(Lf) \, \mathrm{d}\mu.$$

We say that μ satisfies the Poincaré inequality with constant C if for every function f

$$\mathrm{Var}_\mu(f) \leq C\mathcal{E}(f).$$

Theorem 5.3 (Gamma-two criterion)**.** *If* $\mathrm{CD}(\rho, \infty)$ *holds for some positive* ρ *then* μ *satisfies the Poincaré inequality with constant* $1/\rho$*.*

Proof. Recall that letting $\beta(s) = P_s((P_{t-s}f)^2)$ we have $\beta'(s) = 2P_s(\Gamma P_{t-s}f)$. Using the weak commutation we get

$$\beta'(s) \leq \mathrm{e}^{-2\rho(t-s)} P_t(\Gamma f).$$

for every $s \in [0,t]$. Therefore

$$P_t(f^2) - (P_t f)^2 = \beta(t) - \beta(0) \leq \frac{1 - \mathrm{e}^{-2\rho t}}{2\rho} P_t \Gamma(f).$$

Letting t tend to $+\infty$ and using the ergodicity we get the result.

Alternatively, we could define $\alpha(t) = \int P_t(f^2) \mathrm{d}\mu$ and proceed as follows

$$\begin{aligned}\mathrm{Var}_\mu(f) = \alpha(\infty) - \alpha(0) &= \int_0^\infty \alpha(t)\mathrm{d}t = 2\int\int_0^\infty P_t(f)LP_t(f)\mathrm{d}t\mathrm{d}\mu = 2\iint_0^\infty \Gamma(P_t(f))\mathrm{d}t\mathrm{d}\mu \\ &\leq 2\iint_0^\infty \mathrm{e}^{-2\rho t} P_t(\Gamma(f))\mathrm{d}t\mathrm{d}\mu = \Big(\int_0^\infty 2\mathrm{e}^{-2\rho t}\mathrm{d}t\Big)\int \Gamma(f)\mathrm{d}\mu = \frac{1 - \mathrm{e}^{-2\rho t}}{\rho}\int \Gamma(f)\mathrm{d}\mu.\end{aligned}$$

□

The Poincaré inequality is equivalent to an exponential decay of the variance.

Proposition 5.4. *The following are equivalent*

1. *μ satisfies the Poincaré inequality with constant C;*
2. *For any function f, we have* $\mathrm{Var}_\mu(P_t f) \leq \mathrm{e}^{-2t/C} \mathrm{Var}_\mu(f)$.

Proof. Observe that

$$\frac{d}{\mathrm{d}t}\mathrm{Var}(P_t f) = \frac{d}{\mathrm{d}t}\int_M (P_t f)^2 \, \mathrm{d}\mu = -2\mathcal{E}(P_t f).$$

So the direct implication follows from Gronwall and the reverse implication is obtained by differentiating at $t = 0$. □

We close this section by showing that if μ is reversible, the Poincaré inequality is equivalent to an integrated version of the $\mathrm{CD}(\rho,\infty)$ condition.

Proposition 5.5 (Integrated Gamma-two criterion)**.** *Let $\rho > 0$ and consider the following:*

1. *For any f we have $\int_M \Gamma_2(f)\,\mathrm{d}\mu \geq \rho \int_M \Gamma(f)\,\mathrm{d}\mu$;*
2. *μ satisfies Poincaré with constant $1/\rho$.*

We always have 1 ⇒ 2, and if μ is reversible the converse is also true.

Proof. We have seen that $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Var}(P_t f) = -2\mathcal{E}(P_t f)$. By stationarity we have

$$\int_M \Gamma_2(g)\,\mathrm{d}\mu = -\int_M \Gamma(g, Lg)\,\mathrm{d}\mu$$

for all g. Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(P_t f) = 2\int_M \Gamma(P_t f, LP_t f)\,\mathrm{d}\mu = -2\int_M \Gamma_2(P_t f)\,\mathrm{d}\mu.$$

So under the first condition we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(P_t f) \leq -2\rho\,\mathcal{E}(P_t f).$$

Hence by Gronwall $\mathcal{E}(P_t f) \leq \mathrm{e}^{-2\rho t}\mathcal{E}(f)$. Since

$$\mathrm{Var}(f) = \int_0^\infty 2\mathcal{E}(P_t f)\mathrm{d}t$$

we obtain Poincaré with constant $1/\rho$. For the converse inequality, assume (without loss of generality) that f has mean 0. Then using Cauchy–Schwarz and Poincaré we get

$$\begin{aligned}\int_M \Gamma(f)\,\mathrm{d}\mu = -\int_M f(Lf)\,\mathrm{d}\mu &\leq \left(\int_M f^2\,\mathrm{d}\mu\right)^{1/2}\left(\int_M (Lf)^2\mathrm{d}\mu\right)^{1/2} \\ &\leq \left(\frac{1}{\rho}\int_M \Gamma(f)\,\mathrm{d}\mu\right)^{1/2}\left(\int_M (Lf)^2\,\mathrm{d}\mu\right)^{1/2}.\end{aligned}$$

Therefore

$$\int_M (Lf)^2\,\mathrm{d}\mu \geq \rho\int_M \Gamma(f)\,\mathrm{d}\mu.$$

Now reversibility implies that

$$\int_M (Lf)g\,\mathrm{d}\mu = \int_M f(Lg)\,\mathrm{d}\mu = \mathcal{E}(f,g),$$

for any f, g. In particular

$$\int_M (Lf)^2\,\mathrm{d}\mu = \mathcal{E}(f, Lf) = \int_M \Gamma_2(f)\,\mathrm{d}\mu,$$

hence the result. □

5.3 Langevin semigroup

Let $V \colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function and consider the stochastic differential equation

$$\mathrm{d}X_t = \sqrt{2}\,\mathrm{d}B_t - \nabla V(X_t)\,\mathrm{d}t.$$

We assume that it has a unique strong solution, and that this solution does not explode in finite time. This is will be the case under mild assumptions on V but we do not want to spell these out for now. The solution (X_t) is a Markov process and we let (P_t) be the associated semigroup. For $f \in \mathcal{C}_b^2(\mathbb{R}^n)$ we see using Itô's formula that

$$f(X_t) - \int_0^t (\Delta f(X_s) - \langle \nabla V(X_s), \nabla f(X_s)\rangle)\,\mathrm{d}s$$

is a martingale. As a result the operator L given by

$$Lf = \Delta f - \langle \nabla V, \nabla f\rangle$$

is the generator of (P_t). Let μ be the measure given by

$$\mu(dx) = \mathrm{e}^{-V(x)} dx,$$

we claim that μ is reversible for the process. Indeed, if f and g are $\mathcal{C}^2$ smooth and sufficiently decreasing at $+\infty$ then integrating by parts we get

$$\begin{aligned}\int_{\mathbb{R}^n} (\Delta f) g \,\mathrm{d}\mu &= -\int_{\mathbb{R}^n} \langle \nabla f, \nabla (g\mathrm{e}^{-V})\rangle\, dx \\ &= -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle\,\mathrm{d}\mu + \int_{\mathbb{R}^n} \langle \nabla f, \nabla V\rangle\, g\,\mathrm{d}\mu.\end{aligned}$$

Therefore

$$\int_{\mathbb{R}^n} (Lf) g\,\mathrm{d}\mu = -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle\,\mathrm{d}\mu = \int_{\mathbb{R}^n} f(Lg)\,\mathrm{d}\mu,$$

which proves the claim.

If e^{-V} is integrable on $\mathbb{R}^n$ we normalize μ to be a probability measure, and it can be shown that the semigroup (P_t) is ergodic:

$$P_t f \to \int_{\mathbb{R}^n} f\,\mathrm{d}\mu$$

for all f. When the potential $V = |x|^2/2$, the semigroup (P_t) is just the Ornstein–Uhlenbeck semigroup and the measure μ is the Gaussian measure. The Langevin semigroup thus generalizes this construction. The main difficulty of this generalization is that the semigroup (P_t) does not have an explicit expression anymore. However, we shall see that under suitable hypothesis on V, many of the properties that we have established for the Ornstein–Uhlenbeck semigroup remain valid in this context.

Clearly the first order term in L has no effect on the carré du champ, so we still have

$$\Gamma(f,g) = \langle \nabla f, \nabla g\rangle.$$

Let us compute Γ_2:

$$\Gamma_2(f) = \mathrm{Tr}((\mathrm{Hess} f)^2) + \langle \mathrm{Hess} V \nabla f, \nabla f\rangle$$

and $\mathrm{Tr}((\mathrm{Hess} f)^2) = \sum_{i,j} (\partial_{ij} f)^2$ is the squared Hilbert–Schmidt norm of the Hessian of f.

On a Riemannian manifold equipped with the normalized volume measure, the Laplace-Beltrami operator, and a vector field, a new term $\mathrm{Ric}(\nabla,\nabla)$ appears in Γ_2 where Ric stands for the Ricci curvature tensor, making to Bakry – Émery curvature the sum of a curvature of the space and a curvature of the Markov process. A sort of generalization based on Wasserstein – Kantorovich distances, and well suited for discrete space Markov processes, was developped later on by Joulin and Ollivier, while a synthetic notion of curvature was developped by Lott and Villani, Sturm and von Renesse, and the followers.

Lemma 5.6. *Let $\rho \in \mathbb{R}$, the Langevin semigroup satisfies* $\mathrm{CD}(\rho,\infty)$ *if and only if*

$$\mathrm{Hess}V \geq \rho\, I_n \quad \textit{(as quadratic forms, pointwise).}$$

Proof. If $\mathrm{Hess}V \geq \rho\, I_n$ we immediately get $\Gamma_2(f) \geq \rho|\nabla f|^2 = \rho\Gamma(f)$. Conversely, applying $\mathrm{CD}(\rho,\infty)$ to a linear function $f(x) = \langle u, x\rangle$, then $\mathrm{Hess}f = 0$ while $\langle \mathrm{Hess}V u, u\rangle \geq \rho\,|u|^2$. □

Using the main result of the previous section we get the following:

Corollary 5.7. *Let $\rho > 0$ and let μ be a probability measure with density e^{-V} and assume that $\mathrm{Hess}V \geq \rho I_n$ pointwise. Then μ satisfies the following Poincaré inequality*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\rho}\int_{\mathbb{R}^n} |\nabla f|^2 \,\mathrm{d}\mu.$$

5.4 Diffusions

We return to case of a general Markov process (X_t). We have seen how the Γ calculus allows to derive Poincaré. Now we want to do the same thing for log–Sobolev. Given a positive function f we let

$$\beta(s) = P_s\left(\phi(P_{t-s}f)\right),$$

where $\phi = x\log x$. Note that

$$\beta(t) - \beta(0) = P_t(\phi(f)) - \phi(P_t f)$$

evaluated at x is the entropy of f with respect to the measure $\delta_x P_t = P_t(\cdot)(x)$. Then

$$\beta'(s) = P_s\left(L(\phi(g)) - \phi'(g)Lg\right)$$

with $g = P_{t-s}f$. If (X_t) is a Langevin semigroup, as in the previous section, we have

$$L\phi(g) = \Delta(\phi(g)) - \langle \nabla V, \nabla\phi(g)\rangle = \phi'(g)Lg + \phi''(g)|\nabla g|^2.$$

Bearing this example in mind, we say that (X_t) is a diffusion if L satisfies

$$L(\phi(f)) = \phi'(f)Lf + \phi''(f)\Gamma(f),$$

for every f on M and every smooth $\phi\colon \mathbb{R}\to\mathbb{R}$. Note that by taking $\phi(u) = u^2$ we recover the definition of Γ. Typically, a diffusion is the solution of a stochastic differential equation driven by the Brownian motion. On the other hand, a typical example of a Markov that is not a diffusion is a process taking values on a discrete space, like the random walk on a graph. This termonology is not wise, we should say continuous diffusion instead of diffusion since a random walk is a discrete diffusion.

For a diffusion we thus have

$$\beta'(s) = P_s(\phi''(g)\Gamma(g)),$$

which is equal to $\Gamma(g)/g$ when $\phi = x\log x$. The next step is to assume $\mathrm{CD}(\rho,\infty)$ and to use the commutation property

$$\Gamma(g) = \Gamma(P_{t-s}f) \leq \mathrm{e}^{-2\rho(t-s)}\Gamma(P_{t-s}f).$$

It turns out that for log-Sobolev, a stronger commutation is needed.

Proposition 5.8 (strong commutation). *If (X_t) is a diffusion satisfying $\mathrm{CD}(\rho,\infty)$ then*

$$\sqrt{\Gamma P_t} \leq \mathrm{e}^{-\rho t} P_t\left(\sqrt{\Gamma f}\right) \quad \text{for all } f \text{ and } t.$$

Remark. *This is called strong commutation because it implies the weak one immediately by Cauchy–Schwarz. However, at least for diffusions, both commutations are equivalent, since they are both equivalent to $\mathrm{CD}(\rho,\infty)$.*

Proof of Proposition 5.8. Fix $t > 0$ and let

$$\alpha(s) = P_s\left(\sqrt{\Gamma(P_{t-s}f)}\right).$$

If we can prove that $\alpha'(s) \geq \rho\alpha(s)$ for all s then we get $\alpha(t)\mathrm{e}^{-\rho t} \geq \alpha(0)$ by Gronwall, which is the desired inequality. Setting $g = P_{t-s}f$ and using the diffusion property we easily get

$$\alpha'(s) = P_s\left(\Gamma(g)^{-1/2}\Gamma_2(g) - \frac{1}{4}g^{-3/2}\Gamma\Gamma g\right)$$

Therefore

$$\alpha'(s) - \rho\alpha(s) = P_s\left(\Gamma(g)^{-1/2}\left(\Gamma_2(g) - \rho\Gamma(g) - \frac{\Gamma(\Gamma(g))}{4\Gamma(g)}\right)\right)$$

So it is enough to prove that $\Gamma_2(g) \geq \rho\Gamma(g) + \Gamma(\Gamma(g))/4\Gamma(g)$ for all g. This is the purpose of the next lemma. ☐

Lemma 5.9. *For a diffusion satisfying $\mathrm{CD}(\rho,\infty)$ we have*

$$\Gamma_2(f) \geq \rho\Gamma(f) + \frac{\Gamma(\Gamma(f))}{4\Gamma(f)},$$

for all f.

Proof. Let us give the proof for a Langevin diffusion first. As we have seen in the previous section, if a Langevin diffusion satisfies $\mathrm{CD}(\rho,\infty)$ then its V satisfies $\mathrm{Hess}V \geq \rho I_n$, so

$$\begin{aligned}\Gamma_2(f) &= \nabla f(\mathrm{Hess}V)\nabla f + \mathrm{Tr}((\mathrm{Hess}f)^2)\\ &\geq \rho|\nabla f|^2 + \mathrm{Tr}((\mathrm{Hess}f)^2).\end{aligned}$$

Recall that $\Gamma(f) = |\nabla f|^2$ and observe that

$$\begin{aligned}\Gamma(\Gamma(f)) &= |\nabla|\nabla f|^2|^2 = 4|(\mathrm{Hess}f)\nabla f|^2\\ &\leq 4\|\mathrm{Hess}f\|_{\mathrm{op}}^2\,|\nabla f|^2 \leq 4\mathrm{Tr}((\mathrm{Hess}f)^2)\,|\nabla f|^2.\end{aligned}$$

Hence the result in this case. For a general diffusion, we have to proceed using the change of variable for Γ_2, see for instance [24, Lemma 1.3] or [1, Chapter 5]. ☐

5.5 Log–Sobolev inequality for a diffusion

Let (X_t) be a diffusion, we assume that there is a stationary distribution μ which is ergodic. Let f be a positive function and let

$$\beta(s) = P_s(P_{t-s}f \log(P_{t-s}g)),$$

We have seen that

$$\beta'(s) = P_s(\Gamma(P_{t-s}f)/P_{t-s}f).$$

Assuming $\mathrm{CD}(\rho,\infty)$, using strong commutation, and applying Jensen's inequality to the function $(u,v) \mapsto v^2/u$ we get

$$\frac{\Gamma(P_{t-s}f)}{P_{t-s}f} \leq \mathrm{e}^{-2\rho(t-s)} \frac{P_{t-s}(\sqrt{\Gamma(f)})^2}{P_{t-s}f} \leq \mathrm{e}^{-2\rho(t-s)} P_{t-s}\left(\frac{\Gamma(f)}{f}\right).$$

Therefore

$$\beta'(s) \leq \mathrm{e}^{-2\rho(t-s)} P_t\left(\frac{\Gamma(f)}{f}\right).$$

Integrating this between 0 and t we get

$$P_t(f\log f) - P_t f \log(P_t f) \leq \frac{1-\mathrm{e}^{-2\rho t}}{2\rho} P_t\left(\frac{\Gamma(f)}{f}\right).$$

If $\rho > 0$, letting t tend to $+\infty$ and using ergodicity, we obtain the following:

Theorem 5.10. *If a diffusion satisfies* $\mathrm{CD}(\rho,\infty)$ *for some positive* ρ *and has an ergodic stationary measure* μ*. Then* μ *satisfies the follwing logarithmic Sobolev inequality: For any positive* f

$$\mathrm{Ent}_\mu(f) \leq \frac{1}{2\rho}\int_M \frac{\Gamma(f)}{f}\,\mathrm{d}\mu.$$

Remark. *By the diffusion property* $\Gamma(f) = 4f\Gamma(\sqrt{f}) = f\Gamma(f,\log f)$*. So the right-hand side in the log–Sobolev inequality can be rewritten*

$$\int_M \frac{\Gamma(f)}{f}\,\mathrm{d}\mu = \mathcal{E}(f,\log f) = 4\,\mathcal{E}(\sqrt{f}).$$

Corollary 5.11. *Let* μ *be a probability measure on* $\mathbb{R}^n$ *of the form* $\mathrm{d}\mu = \mathrm{e}^{-V}dx$*. If the potential* V *satisfies* $\mathrm{Hess}V \geq \rho I_n$ *for some positive* ρ *then* μ *satisfies the following log–Sobolev inequality: For any positive* f *we have*

$$\mathrm{Ent}_\mu(f) \leq \frac{1}{2\rho}\int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}\,\mathrm{d}\mu.$$

The logarithmic Sobolev inequality is equivalent to an exponential decay of entropy.

Theorem 5.12. *Let* (X_t) *be a diffusion admitting a stationary distribution* μ*. The following are equivalent:*

1. μ *satisfies log–Sobolev with constant* C*: For every positive* f*, we have*

$$\mathrm{Ent}_\mu(f) \leq C\mathcal{E}(f,\log f).$$

2. *For every positive* f *and every time* t *we have*

$$\mathrm{Ent}_\mu(P_t f) \leq \mathrm{e}^{-t/C}\,\mathrm{Ent}_\mu(f).$$

Proof. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Ent}_\mu(P_tf) = \int_M (LP_tf)(\log(P_tf)+1)\,\mathrm{d}\mu = \int_M (LP_tf)\log(P_tf)\,\mathrm{d}\mu.$$

We claim that

$$\int_M (Lg)\log g\,\mathrm{d}\mu = -\mathcal{E}(g,\log g),$$

for all g. Indeed, by stationarity

$$-\mathcal{E}(g,\log g) = \frac{1}{2}\int_M (Lg)\log g\,\mathrm{d}\mu + \frac{1}{2}\int_M g(L\log g)\,\mathrm{d}\mu.$$

Now by the diffusion property $L\log g = g^{-1}Lg - g^{-2}\Gamma(g)$. Therefore, and using stationarity again,

$$\int_M g(L\log g)\,\mathrm{d}\mu = -\int_M \frac{\Gamma(g)}{g}\,\mathrm{d}\mu = -\mathcal{E}(g,\log g),$$

hence the claim. As a result we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Ent}_\mu(P_tf) = -\mathcal{E}(P_tf,\log P_tf).$$

The equivalence is then clear: use Gronwall for one direction and differentiate at $t=0$ for the other. □

As we have seen in the case of the Ornstein–Uhlenbeck semigroup the log-Sobolev inequality is related to hypercontractivity.

Theorem 5.13. *For a diffusion admitting a stationary distribution μ, the following are equivalent:*

1. *μ satisfies the logarithmic Sobolev inequality with constant C;*
2. *For every $p>1$, for every $f\in\mathbb{L}^p(\mu)$ and every $t>0$ we have*

$$\|P_tf\|_{\mathbb{L}^{p(t)}(\mu)} \le \|f\|_{\mathbb{L}^p(\mu)},$$

where $p(t) = 1 + p(\mathrm{e}^{t/C}-1)$.

Proof. Let $f>0$, let $p>1$ and let $p_t = 1+p(\mathrm{e}^{t/C}-1)$. A careful computation shows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\|P_tf\|_{p(t)} = (\int_M f^p\,\mathrm{d}\mu)^{1/p-1}\left(\frac{p'}{p^2}\operatorname{Ent}_\mu(f^p) + \int_M (Lf)f^{p-1}\,\mathrm{d}\mu\right),$$

where $f=P_tf$ and $p=p(t)$ in the right-hand side. Set $g=f^{p/2}$. By the diffusion property

$$Lf = L(g^{2/p}) = \frac{2}{p}g^{2/p-1}Lg + \frac{2}{p}\left(\frac{2}{p}-1\right)g^{2/p-2}\Gamma(g).$$

Thus

$$\begin{aligned}\int_M (Lf)f^{p-1}\,\mathrm{d}\mu &= \frac{2}{p}\int_M g(Lg)\,\mathrm{d}\mu + \frac{2}{p}\left(\frac{2}{p}-1\right)\int_M \Gamma(g)\,\mathrm{d}\mu\\ &= -\frac{4(p-1)}{p^2}\mathcal{E}(g).\end{aligned}$$

So $\frac{\mathrm{d}}{\mathrm{d}t}\|P_tf\|_{p(t)}$ has the same sign as

$$p'(t)\operatorname{Ent}_\mu(g^2) - 4(p-1)\mathcal{E}(g)$$

where $g=(P_tf)^{p/2}$. The equivalence is then straighforward. □

Example (Gaussian Unitary Ensemble)**.** *Let H be a random $N \times N$ Hermitian matrix with density proportional to $\mathrm{e}^{-N\mathrm{Tr}(H^2)}$. It can be show, by a change of variable, that the vector of eigenvalues $(x_1, \ldots, x_N)$ has density in $\mathbb{R}^N$ proportional to*

$$\mathrm{e}^{-N\sum_{k=1}^N x_i^2} \prod_{i<j}(x_i - x_j)^2.$$

If we order the eigenvalues is such a way that say $x_1 \leq \cdots \leq x_N$ then this law μ has density

$$\mathrm{e}^{-H(x_1,\ldots,x_N)} \quad \text{and} \quad H(x_1, \ldots, x_N) = N \sum_{k=1}^N x_k^2 + \sum_{i<j} - \log(x_j - x_i).$$

We have then $\mathrm{Hess}H \geq N$ and therefore, by the Bakry-Émery theory, the probability measure μ satisfies a Poincaré inequality in $\mathbb{R}^N$ with constant $1/N$ and a logarithmic Sobolev inequality with constant $2/N$. This fact is used in [18] for proving local universality. Note that in the Poincaré inequality with constant $1/N$, equality is achieved with $f(x_1, \ldots, x_N) = x_1 + \cdots + x_N$, showing that the constant $1/N$ is optimal.

Example (Ginibre Ensemble)**.** *Let M be a random $N \times N$ matrix with density proportional to $\mathrm{e}^{-N\mathrm{Tr}(MM^*)}$ (same as GUE buth without Hermitianity). It can be show, by a change of variable, that the vector of eigenvalues $(x_1, \ldots, x_N)$ has density in $\mathbb{C}^N$ proportional to*

$$\mathrm{e}^{-N\sum_{k=1}^N |x_i|^2} \prod_{i<j} |x_i - x_j|^2.$$

In contrast with the GUE, this probability measure is not log-concave, regardless of the way we number the eigenvalues. Nevertheless, it can be shown that μ satisfies a Poincaré inequality, but the dependency over N of the constant is not perfectly known.

Finally, let us mention that with the same tools and techniques, it is possible, under $\mathrm{CD}(\rho, \infty)$, to establish, in the diffusion case, for P_t and in particular for P_∞, Φ-Sobolev inequalities with constant $\frac{1-\mathrm{e}^{-2\rho r}}{\rho}$ provided that $(u, v) \mapsto \Phi''(u)v^2$ is convex, as well as Bobkov type I and II Gaussian isoperimetric inequalities under $\mathrm{CD}(\rho, \infty)$.

5.6 Kannan – Lovász – Simonovits conjecture

Let $\mu = \mathrm{e}^{-V}\mathrm{d}x$ with a log-concave probability measure on $\mathbb{R}^n$, in the sense that V is convex, in other words $\mathrm{Hess}(V) \geq 0$ as quadratic forms, pointwise, everywhere. Suppose that the covariance matrix $\Sigma := \mathbb{E}(X \otimes X) - \mathbb{E}(X) \otimes \mathbb{E}(X)$ where $X \sim \mu$, is invertible. The Kannan – Lovász – Simonovits conjecture states that there exists a universal constant $c \geq 1$ (independent of n and μ) such that the optimal Poincaré constant $C_{\mathrm{PI}}(\mu)$ of μ satisfies

$$\|\Sigma\|_{\mathrm{op}} \leq C_{\mathrm{PI}}(\mu) \leq c\|\Sigma\|_{\mathrm{op}}.$$

The conjecture is about the upper bound, the lower bound $\|\Sigma\|_{\mathrm{op}} \leq C_{\mathrm{PI}}(\mu)$ is immediate. Indeed, the Poincaré inequality for μ and the linear function $f = \langle u, \cdot \rangle$, $|u| = 1$, gives

$$\langle \Sigma u, u \rangle = \int f^2 \mathrm{d}\mu - \Big(\int f \mathrm{d}\mu\Big)^2 \leq C_{\mathrm{PI}}(\mu) \int |\nabla f|^2 \mathrm{d}\mu = C_{\mathrm{PI}}(\mu),$$

now since Σ is symmetric positive semidefinite we get $\lambda_{\max}(\Sigma) = \max_{|u|=1}\langle \Sigma u, u \rangle \leq C_{\mathrm{PI}}(\mu)$, and on the the other hand we have $\lambda_{\max}(\Sigma) = \sqrt{\lambda_{\max}(\Sigma^2)} = \max_{|u|=1} |\Sigma u| =: \|\Sigma\|_{\mathrm{op}}$. We refer to [15] for the best results on the KLS conjecture at the time of writing these notes.

A result stronger than the Bakry – Émery criterion and weaker than the Kannan – Lovász – Simonovits conjecture is known as the Bascamp – Lieb inquality. It states that if $\mu = \mathrm{e}^{-V}\mathrm{d}x$ with V strictly convex in the sense that $\mathrm{Hess}(V) > 0$ as quadratic forms, pointwise, everywhere, then for all test function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int f^2 \mathrm{d}\mu - \Big(\int f \mathrm{d}\mu\Big)^2 \leq \int \langle \mathrm{Hess}(V)^{-1} \nabla f, \nabla f \rangle \mathrm{d}\mu.$$

It can be proved by induction over the dimension, and integration by parts for $n = 1$.

Chapter 6

Brenier and Caffarelli theorems

In this section we give an alternate proof of the results of the previous section based on optimal transport techniques. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ and μ be a probability measure on $\mathbb{R}^n$. The *pushforward* of μ by T is the measure ν given, for every Borel subset A of $\mathbb{R}^n$, by

$$\nu(A) = \mu(T^{-1}(A)).$$

In other words ν is the law of $T(X)$ where X is a random vector having law μ, and thus

$$\int_{\mathbb{R}^n} h\,\mathrm{d}\nu = \int_{\mathbb{R}^n} h \circ T\,\mathrm{d}\mu$$

for every test function h.

6.1 Brenier theorem

In dimension 1, if μ and ν are two probability measures on $\mathbb{R}$ with cumulative distribution functions $F = \mu((-\infty, \bullet])$ and $G = \nu((-\infty, \bullet])$ respectively, then the map $F^{-1} \circ G$ is non-decreasing and pushes forward μ to ν. The Brenier theorem states that in arbitrary dimension, it is still possible to push forward using the gradient of a convex function, which can be seen as a multivariate generalization of the notion of non-decreasing function.

Theorem 6.1 (Brenier)**.** *If μ and ν are two probability measures on $\mathbb{R}^n$ with μ absolutely continuous with respect to the Lebesgue measure then there exists a unique map $T : \mathbb{R}^n \to \mathbb{R}^n$ pushing forward μ to ν and $T = \nabla\phi$ with ϕ convex.*

The uniqueness of the map T must be understood almost everywhere.

The convex function ϕ is obviously not unique but its gradient is unique.

By Rademacher's theorem a convex function is differentiable almost everywhere in its domain, so if μ is absolutely continuous and if the support of μ is contained in the domain of ϕ then $\nabla\phi$ makes sense almost everywhere for μ.

Remark (Relation to Wasserstein−Kantorovich distance)**.** *If μ and ν have finite second moment and if $T = \nabla\phi$ is the Brenier map pushing forward μ to ν then*

$$W_2(\mu, \nu)^2 = \min_{\pi \in \Pi(\mu,\nu)} \int \frac{|x - y|^2}{2} \pi(\mathrm{d}x, \mathrm{d}y) = \int \frac{|x - T(x)|^2}{2} \mathrm{d}\mu(\mathrm{d}x).$$

In other words the optimal coupling is deterministic: $\pi(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x)\delta_{T(x)}(\mathrm{d}y)$.
The transport map $T = \nabla\phi$ realizes an optimal transport *of μ to ν.*
Actually the optimality in the dual formulation of W_2 leads to a proof of Brenier's theorem.

Remark (Reverse Brenier map, Legendre transform, convex duality). *If ν is absolutely continuous with respect to the Lebesgue measure then $\nabla\phi$ is invertible and $(\nabla\phi)^{-1} = \nabla\phi^*$ is the Brenier map between ν and μ, where*

$$\phi^*(y) = \sup_x \{\langle x, y\rangle - \phi(x)\}.$$

is the Legendre transform of ϕ (it is the also the gradient of a convex function).

Let us discuss the regularity of the Brenier map $\nabla\phi$. Clearly $\nabla\phi$ need not be continuous. For instance if μ is uniform on $[0,1]$ and ν is uniform on $[0,1/2]\cup[3/2,2]$ then the Brenier map must be the identity on $[0,1/2[$ and identity plus 1 on $]1/2,1]$. It turns out that the correct hypothesis for the regularity of the Brenier map is convexity of the support of the target measure. We state below a theorem that follows from the work of Caffarelli,

Theorem 6.2 (Caffarelli). *Assume that μ and ν are absolutely continuous, that their respective supports K and L are convex, and that their respective densities f, g are bounded away from 0 and $+\infty$ on K and L respectively. Then the Brenier map $\nabla\phi$ is an homeomorphism between the interior of K and that of L. Moreover if f and g are continuous then $\nabla\phi$ is a $\mathcal{C}^1$–diffeomorphism.*

6.2 Monge – Ampère equation

When $\nabla\phi$ is a $\mathcal{C}^1$–diffeomorphism we can apply the change of variable formula $y = \phi(x)$, namely for every test function h, since $\mathrm{Jac}\nabla\phi = \nabla^2\phi$ (Hessian),

$$\int_L h(y)g(y)\,dy = \int_K h\left(\nabla\phi(x)\right)\, g\left(\nabla\phi(x)\right)\, \det(\nabla^2\phi(x))\,dx.$$

On the other hand, by definition of the Brenier map

$$\int_L h(y)g(y)\,dy = \int_{\mathbb{R}^n} h\,\mathrm{d}\nu = \int_{\mathbb{R}^n} h\circ\nabla\phi\,\mathrm{d}\mu = \int_K h\left(\nabla\phi(x)\right) f(x)\,dx.$$

Since this is valid for every test function h we obtain the following equality

$$g\left(\nabla\phi(x)\right)\,\det(\nabla^2\phi(x)) = f(x), \tag{6.1}$$

for every x in the interior of K. This is called Monge–Ampère equation.

Remark (From Monge – Ampère to Poisson – Langevin). *When $\phi(x) = \frac{1}{2}|x|^2$ the Monge – Ampère simply reads $g(x) = f(x)$. Let us consider a perturbation or linearization around this case by taking $\phi(x) = \frac{1}{2}|x|^2 + \varepsilon\psi(x) + O(\varepsilon^2)$ and $g(x) = (1+\varepsilon h(x) + O(\varepsilon^2))f(x)$, then, as $\varepsilon\to 0$, we find the Poisson equation for the Langevin operator:*

$$\left(-\Delta - \frac{\nabla f}{f}\cdot\nabla\right)\psi = h.$$

In other words, this reads $-(\Delta - \nabla V\cdot\nabla)\psi = h$ if we write $f = \mathrm{e}^{-V}$. In the same spirit, the Wasserstein – Kantorivich distance can be interpreted as an inverse Sobolev norm.

6.3 Caffarelli contraction theorem

Let us keep the notion just above. Here is the main result of this section.

Theorem 6.3 (Caffarelli's contraction theorem). *Let $\alpha, \beta > 0$, let μ, ν be two probability measures on $\mathbb{R}^n$ of the form*

$$\mathrm{d}\mu = \mathrm{e}^{-V}\, dx, \quad \mathrm{d}\nu = \mathrm{e}^{-W}\, dx,$$

and assume that the potentials V and W are smooth and satisfy, pointwise on $\mathbb{R}^n$,

$$\nabla^2 V \leq \alpha I_n, \quad \nabla^2 W \geq \beta I_n.$$

Then the Brenier map $\nabla\phi$ between μ and ν is Lipschitz with constant $\sqrt{\alpha/\beta}$.

Proof. We only give a formal proof of the result, based on a maximum principle.

Let us begin with the undimensional case $n = 1$ to focus on the idea without the multivariate complexity. Taking the logarithm in the Monge−Ampère equation (6.1) gives $\frac{1}{2}\log(\varphi''^2) = \log|\varphi''| = -V + W(\varphi')$, and taking the derivative twice gives

$$\frac{\varphi''''\varphi'' - \varphi'''^2}{\varphi''^2} = -V'' + W''(\varphi')\varphi''^2 + W'(\varphi')\varphi'''.$$

Now if φ'' has a maximum at $x = x_*$ then $\varphi'''(x_*) = 0$ and $\varphi''''(x_*) \leq 0$, and thus

$$0 \geq -V''(x_*) + W''(\varphi'(x_*))\varphi''^2(x_*) \quad \text{hence} \quad \varphi''^2(x_*) \leq \alpha/\beta.$$

The argument is attractive but a maximum at the boundary may produce difficulties.

Let us consider now the multidimensional case $n \geq 1$. Observe first that the Lipschitz constant of $\nabla\phi$ is the supremum of the operator norm of $\nabla^2\phi$. So it is enough to prove $\|\nabla^2\phi(x)\|_{\mathrm{op}} \leq \sqrt{\alpha/\beta}$ for every x. Besides since ϕ is convex $\nabla^2\phi$ is a positive matrix so this amounts to proving that $\langle\nabla^2\phi(x)u, u\rangle \leq \sqrt{\alpha/\beta}$ for every unit vector u and every $x \in \mathbb{R}^n$. Now we fix a direction u and we assume that the map

$$\ell \colon x \mapsto \langle\nabla^2\phi(x)u, u\rangle$$

attains its maximum at some point $x_0 \in \mathbb{R}^n$. Taking the logarithm of the Monge–Ampère equation (6.1) we obtain in this case

$$\log\det\Big(\nabla^2\phi(x)\Big) = -V(x) + W\left(\nabla\phi(x)\right).$$

Now we want to differentiate this equation twice in the direction u. To differentiate the left hand side, observe that if A is an invertible matrix

$$\log\det(A + H) = \log\det(A) + \mathrm{tr}(A^{-1}H) + o(H);$$
$$(A + H)^{-1} = A^{-1} - A^{-1}HA^{-1} + o(H).$$

We obtain (omitting variables)

$$\begin{aligned} &- \mathrm{tr}\Big((\nabla^2\phi)^{-1}(\partial_u\nabla^2\phi)(\nabla^2\phi)^{-1}(\partial_u\nabla^2\phi)\Big) + \mathrm{tr}\Big((\nabla^2\phi)^{-1}\partial_{uu}\nabla^2\phi\Big) \\ &= -\partial_{uu}V + \sum_i \partial_i W \partial_{iuu}\phi + \sum_{ij} \partial_{ij}W(\partial_{iu}\phi)(\partial_{ju}\phi). \end{aligned} \tag{6.2}$$

We shall use this equation at x_0. We claim that

$$\operatorname{tr}\left((\nabla^2\phi)^{-1}(\partial_u\nabla^2\phi)(\nabla^2\phi)^{-1}(\partial_u\nabla^2\phi)\right) \geq 0.$$

Indeed, $\nabla^2\phi$ is positive so $(\nabla^2\phi)^{-1}$ is also positive and since $\partial_u\nabla^2\phi$ is symmetric, we get

$$(\partial_u\nabla^2\phi)(\nabla^2\phi)^{-1}(\partial_u\nabla^2\phi) \geq 0.$$

Since the product of two positive matrices has positive trace[1] we get the claim. Since function ℓ attains its maximum at x_0 we have $\nabla^2\ell(x_0) \leq 0$. Therefore

$$\operatorname{tr}\left((\nabla^2\phi)^{-1}\partial_{uu}\nabla^2\phi\right) = \operatorname{tr}\left((\nabla^2\phi)^{-1}\nabla^2\ell\right) \leq 0.$$

In the same way

$$\sum_i \partial_i W \partial_{iuu}\phi = \langle \nabla W, \nabla \ell\rangle = 0.$$

So at point x_0, equality (6.2) gives

$$\sum_{ij} \partial_{ij} W (\partial_{iu}\phi)(\partial_{ju}\phi) \leq \partial_{uu} V.$$

Now the hypothesis made on V and W give $\partial_{uu}V \leq \alpha$ and

$$\sum_{ij} \partial_{ij} W (\partial_{iu}\phi)(\partial_{ju}\phi) \geq \beta \sum_i (\partial_{iu}\phi)^2 = \beta|\nabla^2\phi(u)|^2.$$

Since u has norm 1, we get

$$\ell(x_0) = \langle \nabla^2\phi(x_0)u, u\rangle \leq |\nabla^2\phi(x_0)(u)| \leq \sqrt{\frac{\alpha}{\beta}}.$$

Therefore $\ell(x) \leq \sqrt{\alpha/\beta}$ for every x which is the desired inequality. $\square$

Now we give an example of an application of the previous result. Let γ_n be the standard Gaussian measure and let μ be a probability measure satisfying $\mathrm{d}\mu = \mathrm{e}^{-V}\,dx$ with $\nabla^2 V \geq \rho I_n$ for some positive ρ (strong log-concavity, as in the chapter on Bakry – Émery). According to Caffarelli's contraction theorem the Brenier map between γ_n and μ is Lipschitz with constant $\sqrt{1/\rho}$. Let us derive the Poincaré inequality for μ. Let $\nabla\phi$ be the Brenier map from γ_n to μ. Since $\nabla\phi$ pushes forward γ_n to μ

$$\int_{\mathbb{R}^n} f \,\mathrm{d}\mu = \int_{\mathbb{R}^n} f \circ \nabla\phi \,\mathrm{d}\gamma_n$$

for every test function f. The same holds for f^2 and we get

$$\operatorname{Var}_\mu(f) = \operatorname{Var}_{\gamma_n}(f \circ \nabla\phi).$$

Applying Poincaré inequality for the Gaussian measure we get

$$\operatorname{Var}_{\gamma_n}(f \circ \nabla\phi) \leq \int_{\mathbb{R}^n} |\nabla(f\circ\nabla\phi)|^2 \,\mathrm{d}\gamma_n.$$

Now since $\nabla\phi$ is Lipschitz with constant $\sqrt{1/\rho}$

$$|\nabla(f\circ\nabla\phi)|^2 = |\nabla^2\phi\,(\nabla f\circ\nabla\phi)\,|^2 \leq \|\nabla^2\phi\|_{\mathrm{op}}\,|\nabla f\circ\nabla\phi|^2 \leq \frac{1}{\rho}\,|\nabla f\circ\nabla\phi|^2.$$

[1] If A and B are $n\times n$ real symmetric positive semidefinite then $\operatorname{Tr}(AB) = \operatorname{Tr}(\sqrt{A}\sqrt{B}(\sqrt{A}\sqrt{B})^\top) \geq 0$.

Therefore

$$\int_{\mathbb{R}^n} |\nabla(f\circ\nabla\phi)|^2\,\mathrm{d}\gamma_n \le \frac{1}{\rho}\int_{\mathbb{R}^n} |\nabla f\circ\nabla\phi|^2\,\mathrm{d}\gamma_n = \frac{1}{\rho}\int_{\mathbb{R}^n} |\nabla f|^2\,\mathrm{d}\mu.$$

So μ satisfies Poincaré with constant $1/\rho$. A very similar argument shows that μ satisfies log–Sobolev with constant $1/(2\rho)$.

To sum up, we have seen in this section that a measure having a uniformly convex potential is a Lipschitz image of the Gaussian measure. Since the pushfoward by a Lipschitz map preserves log-Sobolev and Poincaré this shows that such a measure satisfies log-Sobolev and Poincaré. This is a remarkable alternative to the Γ_2 or curvature-dimension Bakry – Émery criterion, however its extension to manifolds is unclear.

6.4 From Monge – Ampère to Gaussian log-Sobolev

Following [16], let us give a proof of the optimal logarithmic Sobolev inequality for the standard Gaussian measure γ_n by using directly the Monge – Ampère equation. Let $f:\mathbb{R}^n\to\mathbb{R}_+$ be such that $\int f\mathrm{d}\gamma_n = 1$. Let $T=\nabla\phi$ be the Brenier map pushing forward $f\mathrm{d}\gamma_n$ to γ_n. We set $\theta(x) := \phi(x) - \frac{1}{2}|x|^2$ so that $\nabla\phi(x) = x + \nabla\theta(x)$. We have $\mathrm{Hess}(\theta)(x) + I_n \ge 0$, and the Monge – Ampère equation gives

$$f(x)\mathrm{e}^{-\frac{|x|^2}{2}} = \det(I_n + \mathrm{Hess}(\theta)(x))\mathrm{e}^{-\frac{|x+\nabla\theta(x)|^2}{2}}.$$

Taking the logarithm gives

$$\begin{aligned}\log f(x) &= -\frac{|x+\nabla\theta(x)|^2}{2} + \frac{|x|^2}{2} + \log\det(I_n + \mathrm{Hess}(\theta)(x))\\ &= -x\cdot\nabla\theta(x) - \frac{|\nabla\theta(x)|^2}{2} + \log\det(I_n + \mathrm{Hess}(\theta)(x))\\ &\le -x\cdot\nabla\theta(x) - \frac{|\nabla\theta(x)|^2}{2} + \Delta\theta(x),\end{aligned}$$

where we have used $\log(1+t)\le t$ for $1+t>0$ and the eigenvalues of the positive symmetric matrix $I_n + \mathrm{Hess}(\theta)(x)$. Now integration with respect to $f\mathrm{d}\gamma_n$ gives

$$\int f\log f\mathrm{d}\gamma_n \le \int f(\Delta\theta - x\cdot\nabla\theta)\mathrm{d}\gamma_n - \int\frac{|\nabla\theta|^2}{2}f\mathrm{d}\gamma_n.$$

Finally, using integration by parts ($\Delta\theta - x\cdot\nabla\theta$ is O.-U.!), we get

$$\begin{aligned}\int f\log f\mathrm{d}\gamma_n &\le -\int\frac{1}{2}\Big|\sqrt{f}\nabla\theta + \frac{\nabla f}{\sqrt{f}}\Big|^2\mathrm{d}\gamma_n + \frac{1}{2}\int\frac{|\nabla f|^2}{f}\mathrm{d}\gamma_n\\ &\le \frac{1}{2}\int\frac{|\nabla f|^2}{f}\mathrm{d}\gamma_n.\end{aligned}$$

Recall that $T=\nabla\phi = x+\nabla\theta$ pushes forward ν to γ_n, where $\mathrm{d}\nu = f\mathrm{d}\gamma_n$. Therefore

$$\int\frac{|\nabla\theta|^2}{2}f\mathrm{d}\gamma_n = \int\frac{|x-T(x)|^2}{2}\mathrm{d}\nu = W_2^2(\nu,\gamma_n).$$

Actually, beyond the log-Sobolev inequality and beyond the Gaussian measure, it is possible to obtain by this way, from the Monge – Ampère equation, HWI[2] functional inequalities for strongly log-concave measures, see for instance [3, Section 9.3]. From this point of view, optimal transportation provides a partial alternative to the Bakry – Émery on $\mathbb{R}^n$.

[2]Here H stands for entropy, W for Wasserstein, and I for Fisher information.

Chapter 7

Discrete space

7.1 Bernoulli distributions

For any $p \in (0,1)$, we consider the Bernoulli probability measure

$$\mathrm{Ber}(p) := p\delta_1 + q\delta_0$$

on the two-points space $\{0,1\}$, with $q := 1-p$. For any $f : \{0,1\} \to \mathbb{R}$ we define

$$(\mathrm{D}f)(x) = f(x+1) - f(x)$$

where $1+1=0$ on $\{0,1\} = \mathbb{Z}/2\mathbb{Z}$, and thus $(\mathrm{D}f)(1) = -(\mathrm{D}f)(0) = f(1) - f(0)$. In this way D is a forward difference operator on the discrete circle $\mathbb{Z}/2\mathbb{Z} = \{0,1\}$.

Theorem 7.1 (Poincaré equality for Bernoulli laws). *For any $f : \{0,1\} \to \mathbb{R}$,*

$$\mathrm{Var}_{\mathrm{Ber}(p)}(f) = pq\mathbb{E}_{\mathrm{Ber}(p)}(\mathrm{D}f)^2 \tag{7.1}$$

Proof. For every $f : \{0,1\} \to \mathbb{R}$, the function $(\mathrm{D}f)^2 : \{0,1\} \to \mathbb{R}$ is constant and equal to $(f(1) - f(0))^2$. The result follows since $\mathrm{Var}_{\mathrm{Ber}(p)}f = pq(f(1) - f(0))^2$. □

Theorem 7.2 (Logarithmic Sobolev inequality for Bernoulli laws). *For any $f : \{0,1\} \to \mathbb{R}$,*

$$\mathrm{Ent}_{\mathrm{Ber}(p)}(f^2) \leq c_p pq\mathbb{E}_{\mathrm{Ber}(p)}((\mathrm{D}f)^2) \tag{7.2}$$

where

$$c_p := \begin{cases} \dfrac{\log(q) - \log(p)}{q-p} & \text{if } p \neq 1/2, \\ 2 & \text{if } p = 1/2. \end{cases}$$

Moreover, equality is achieved when $pf(1) = qf(0)$.

The function $p \in (0,1) \mapsto c_p$ is convex, symmetric with respect to the vertical axis of equation $p = 1/2$, tends to $+\infty$ when p tends to 0^+ or 1^-, and achieves its minimum $2 = c_{1/2} = \lim_{p\to1/2} c_p$ for $p = 1/2$.

Proof. By scaling we can assume that $f(0) = 1$. Set $(f(0)^2, f(1)^2) = (1, u^2)$. If $p = q = 1/2$, the desired result is nothing else but the following valid bound

$$u^2\log(u^2) - (1+u^2)\log\frac{1+u^2}{2} \leq (u-1)^2, \quad u \geq 0.$$

This is actually equivalent to what we did sligtly differently for the CLT proof of the logarithmic Sobolev inequality for the Gauss distribution.

Suppose now that $p \neq q$. By symmetry we can assume that say $p < q$. Moreover, since $|\mathrm{D}|f|| \leq |\mathrm{D}f|$, we can further assume that $f \geq 0$. Now set

$$\psi(u) := \mathrm{Ent}_{\mathrm{Ber}(p)}(f^2) = pu^2 \log(u^2) - (q + pu^2) \log(q + pu^2), \quad u \geq 0.$$

Our objective is to compute $c := \sup_{u \geq 0}(u-1)^{-2}\psi(u)$. The critical values of u are solutions of $(u-1)\psi'(u) = 2\psi(u)$, and q/p is a critical value. Now elementary computations reveal that $\psi(1) = \psi'(1) = 0$, that ψ is convex, and that ψ'' achieves its maximum when $u = q/p$. An elementary study shows that c is finite, and is achieved for a unique value of u which belongs to $(0,1) \cup (1,\infty)$, which is thus necessarily the critical value q/p, which gives $c = (q/p-1)^{-2}\psi(q/p)$. □

The bound $|\mathrm{D}|f|| \leq |\mathrm{D}f|$ allows to reduce the logarithmic Sobolev inequality to the case where f is non negative, in other words that for all $f : \{0,1\} \to [0,\infty)$,

$$\mathrm{Ent}_{\mathrm{Ber}(p)}(f) \leq \frac{\log(q) - \log(p)}{q-p} pq\mathbb{E}_{\mathrm{Ber}(p)}((\mathrm{D}\sqrt{f})^2).$$

This is an $\mathbb{L}^1$ version of the logarithmic Sobolev inequality. The following theorem provides many other $\mathbb{L}^1$ inequalities, which are not equivalent due to the lack of chain rule $\nabla\alpha(f) = \alpha'(f)\nabla f$ for the discrete gradient D. Namely, what is lacking here is the following, valid for any smooth $f : \mathbb{R}^n \to \mathbb{R}$:

$$4(\nabla\sqrt{f})^2 = \frac{|\nabla f|^2}{f} = \langle \nabla f, \log(f) \rangle. \tag{7.3}$$

For any $u, v \in \mathbb{R}$ with $u \geq 0$ and $u + v \geq 0$, define $\phi(u) := u\log(u)$ and

$$\begin{aligned} A(u,v) &:= \phi(u+v) - \phi(u) - \phi'(u)v, \\ B(u,v) &:= (\phi'(u+v) - \phi'(u))v, \\ C(u,v) &:= \phi''(u)v^2. \end{aligned}$$

Then it turns out that A, B, C are non-negative and convex, and

$$\begin{aligned} A(u,v) &= (u+v)(\log(u+v) - \log(u)) - v, \\ B(u,v) &= (\log(u+v) - \log(u))v, \\ C(u,v) &= \frac{v^2}{u}. \end{aligned}$$

Moreover for any $u \in [0,\infty)$ and $v \in \mathbb{R}$ with $u + v \in [0,\infty)$,

$$A(u,v) \leq B(u,v) \quad \text{and} \quad A(u,v) \leq C(u,v). \tag{7.4}$$

Note that $B(u,v) \leq C(u,v)$ if $v \geq 0$. Furthermore for any function $f > 0$,

$$\begin{aligned} A(f,\mathrm{D}f) &= (f + \mathrm{D}f)\mathrm{D}\log(f) - \mathrm{D}f, \\ B(f,\mathrm{D}f) &= \mathrm{D}f\mathrm{D}\log(f), \\ C(f,\mathrm{D}f) &= \frac{(\mathrm{D}f)^2}{f}. \end{aligned} \tag{7.5}$$

The following theorem states that these three expressions can serve as a right hand side of an entropic inequality called “modified logarithmic Sobolev inequalities”.

Theorem 7.3 (Modified logarithmic Sobolev inequalities for Bernoulli laws)**.** *Let A, B, C be as in* (7.4)*. Then for any $f : \{0,1\} \to (0,\infty)$,*

$$\mathrm{Ent}_{\mathrm{Ber}(p)}(f) \leq pq\mathbb{E}_{\mathrm{Ber}(p)}(A(f,\mathrm{D}f)), \tag{7.6}$$

and in particular

$$\mathrm{Ent}_{\mathrm{Ber}(p)}(f) \leq pq\mathbb{E}_{\mathrm{Ber}(p)}(B(f,\mathrm{D}f)), \tag{7.7}$$

and

$$\mathrm{Ent}_{\mathrm{Ber}(p)}(f) \leq pq\mathbb{E}_{\mathrm{Ber}(p)}(C(f,\mathrm{D}f)). \tag{7.8}$$

Proof. First by using the comparisons (7.4) and the formulas (7.5), we deduce immediately (7.7) and (7.8) from (7.6). Next, to prove (7.6), fix f and define

$$U : [0,1] \to \mathbb{R} \quad \text{by} \quad U(p) := \mathrm{Ent}_{\mathrm{Ber}(p)} f - pq\mathbb{E}_{\mathrm{Ber}(p)}(A(f,\mathrm{D}f))$$

Set $(a,b) := (f(0), f(1))$. We have

$$U(p) = q\phi(a) + p\phi(b) - \phi(qa+pb) - pq(qA(f,\mathrm{D}f)(0) + pA(f,\mathrm{D}f)(1))$$

where $\phi(x) := x\log(x)$. We use an argument due to Sergey Bobkov, see [23]. It suffices to show that $U \leq 0$ on $[0,1]$. To kill the polynomial terms (with respect to p) in $U(p)$, let us compute the fourth derivative of U with respect to p, namely

$$U''''(p) = -(b-a)^4\phi''''(qa+pb).$$

Since ϕ'' is convex, we have $U'''' \leq 0$ on $(0,1)$ and thus U'' is concave. Consequently, there exists $0 \leq p_0 \leq p_1 \leq 1$ such that $U'' \leq 0$ on $[0,p_0] \cup [p_1,1]$ and $U'' \geq 0$ on $[p_0,p_1]$. Hence, U is concave on $[0,p_0]$. But $U(0) = 0$ and $U'(0) \leq 0$, et thus $U \leq 0$ on $[0,p_0]$ by concavity. It follows that $U(p_0) \leq 0$. By symmetry $U \leq 0$ on $[p_1,1]$ and $U(p_1) \leq 0$. Now since U is convex on $[p_0,p_1]$ and negative at the boundaries, it is also negative on $[p_0,p_1]$. It follows that $U \leq 0$ on $[0,1]$, and (7.6) is proved. Actually the proof works also for (7.7) and (7.8). □

7.2 Poisson distributions

For any real parameter $\lambda \geq 0$, the Poisson probability measure of parameter λ is

$$\mathrm{Poi}(\lambda) := \mathrm{e}^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\delta_n.$$

Let D be the finite difference operator defined for any $f : \mathbb{N} \to \mathbb{R}$ by

$$(\mathrm{D}f)(n) := f(n+1) - f(n), \quad n \in \mathbb{N}.$$

Theorem 7.4 (Poincaré inequality for Poisson laws)**.** *For any $f : \mathbb{N} \to \mathbb{R}$,*

$$\mathrm{Var}_{\mathrm{Poi}(\lambda)}(f) \leq \lambda\mathbb{E}_{\mathrm{Poi}(\lambda)}((\mathrm{D}f)^2). \tag{7.9}$$

Equality is achieved when f is affine and the constant λ is thus optimal.

Proof. We may assume that f is bounded by using a cutoff. Fix a bounded $f : \mathbb{N} \to \mathbb{R}$. Define $s_n : \{0,1\}^n \to \mathbb{N}$ by $s_n(x) := x_1 + \cdots + x_n$, $x \in \{0,1\}^n$. Set $F_n := f \circ s_n : \{0,1\}^n \to \mathbb{R}$. Using the Poincaré equality (7.1) and the tensorization inequality (1.10) for the product $(\{0,1\}^n, \mathrm{Ber}(p)^{\otimes n})$, we get

$$\mathrm{Var}_{\mathrm{Ber}(p)^{\otimes n}}(F) \leq p(1-p)\mathbb{E}_{\mathrm{Ber}(p)^{\otimes n}}\Big(\sum_{i=1}^n (\mathrm{D}_i F)^2\Big)$$

where D_i is the binary operator D acting on the i-th coordinate. Now $s_n(x)$ and $n - s_n(x)$ count the number of 1's and 0's in $x \in \{0,1\}^n$, and it follows then that

$$\sum_{i=1}^n (\mathrm{D}_i F_n)^2(x) = (n - s_n(x))(\mathrm{D}f)^2(s_n(x)) + s_n(x)(\mathrm{D}^* f)^2(s_n(x)), \quad x \in \{0,1\}^n,$$

where D^* is the backward difference operator defined by

$$(\mathrm{D}^* f)(s) := f(s-1) - f(s), \quad s \in \mathbb{N} \setminus \{0\}.$$

Now the law of s_n under $\mathrm{Ber}(p)^{\otimes n}$ is the binomial law

$$\mathrm{Bin}(n,p) = \sum_{k=0}^n \binom{k}{n} p^k q^{n-k} \delta_k.$$

Therefore, setting $G(s) := -s(\mathrm{D}f)^2(s) + s(\mathrm{D}^* f)^2(s)$, we get, with $S_n \sim \mathrm{Bin}(n,p)$,

$$\mathrm{Var}(F(S_n)) \leq np(1-p)\mathbb{E}((\mathrm{D}f)^2(S_n)) + p(1-p)\mathbb{E}(G(S_n)).$$

Set $p = p_n$ such that $\lim_{n\to\infty} np_n = \lambda$, for example take $p_n = \lambda/n$. Then $np_n(1-p_n) \to \lambda$, $p_n(1-p_n) \to 0$, and the law of small numbers states that

$$\lim_{n\to\infty} \mathrm{Bin}(n, p_n) = \mathrm{Poi}(\lambda)$$

at least in the sense of weak convergence with respect to Lipschitz test functions $\mathbb{N} \to \mathbb{R}$. Now G is Lipschitz since f is bounded and we get, with $S_\infty \sim \mathrm{Poi}(\lambda)$,

$$\mathrm{Var}(f(S_\infty)) \leq \lambda \mathbb{E}((\mathrm{D}f)^2(S_\infty)).$$

□

Let us say that the tail of a probability measure μ on $\mathbb{N}$ is sub-exponential (respectively sub-Gaussian) when for $X \sim \mu$, some constants $c, C > 0$, and any $r \geq 0$, we have $\mathbb{P}(X \geq r) \leq C\mathrm{e}^{-cr}$ (respectively $\mathbb{P}(X \geq r) \leq C\mathrm{e}^{-cr^2}$). It turns out that the tail of a Poisson law is sub-exponential but is not sub-Gaussian.

Lemma 7.5 (Poisson tail)**.** *For any $\lambda > 0$, there exists explicit affine functions $\ell_\pm : \mathbb{N} \to \mathbb{R}$ such that if $X \sim \mathrm{Poi}(\lambda)$ then for any $r \in \mathbb{N}$,*

$$\mathrm{e}^{-r\log(r) + \ell_-(r)} \leq \mathbb{P}(X \geq r) \leq \mathrm{e}^{-r\log(r) + \ell_+(r)}.$$

Proof. If $X \sim \mathrm{Poi}(\lambda)$ with $\lambda > 0$, then for any $r \in \mathbb{N}$,

$$\mathbb{P}(X \geq r) = \mathrm{e}^{-\lambda} \sum_{k=r}^\infty \frac{\lambda^k}{k!} \begin{cases} \leq & \mathrm{e}^{-\lambda} \frac{\lambda^r}{r!} \sum_{k=r}^\infty \frac{\lambda^{k-r}}{(k-r)!} = \frac{\lambda^r}{r!}, \\ \geq & \mathrm{e}^{-\lambda} \frac{\lambda^r}{r!}, \end{cases}$$

and it remains to use the Stirling bound $\sqrt{2\pi}r^{r+\frac{1}{2}}\mathrm{e}^{-r} \le r! \le \mathrm{e}r^{r+\frac{1}{2}}\mathrm{e}^{-r}$.

Note that the Poisson tail can be remarkably expressed with the incomplete Gamma function, namely, integration by parts gives, for any integer $r \ge 1$,

$$\mathbb{P}(X \ge r) = \frac{1}{(r-1)!}\int_0^\lambda t^{r-1}\mathrm{e}^{-t}\,\mathrm{d}t,$$

which allows an asymptotic expansion as $r \to \infty$ via the Laplace method. □

The Poisson tail is incompatible with a logarithmic Sobolev inequality.

Theorem 7.6 (Lack of logarithmic Sobolev inequality for Poisson laws)**.** *For any $\lambda > 0$, the Poisson law* $\mathrm{Poi}(\lambda)$ *does not satisfy to a logarithmic Sobolev inequality: there is not any constant $c > 0$ such that for any bounded $f : \mathbb{N} \to \mathbb{R}$,*

$$\mathrm{Ent}_{\mathrm{Poi}(\lambda)}(f^2) \le c\mathbb{E}_{\mathrm{Poi}(\lambda)}((\mathrm{D}f)^2).$$

Proof. We proceed by contradiction. Suppose that the logarithmic Sobolev inequality holds for some $c > 0$. Then, for any $r \in \mathbb{N}$, this inequality used with the indicator $f = \mathbf{1}_{A_r}$ of the infinite set $A_r := \{r+1, \ldots\} = \mathbb{N} \cap (r, \infty)$ yields

$$-\mathbb{P}(X > r)\log\mathbb{P}(X > r) \le c\mathbb{P}(X = r),$$

an inequality which contradicts the finiteness of c when $r \to \infty$.

Alternatively, we can use the Herbst argument in order to deduce from the logarithmic Sobolev inequality that the tail is sub-Gaussian. The lack of chain rule for the discrete gradient D can be circumvented as in the proof of Theorem 7.8. In contrast, the Poincaré inequality (7.9) implies via the Herbst argument a sub-exponential tail, which is fairly compatible with the Poisson law. □

Theorem 7.7 (Modified logarithmic Sobolev inequalities for Poisson laws)**.** *Let A, B, C be as in* (7.5)*. For any $f : \mathbb{N} \to (0, \infty)$,*

$$\mathrm{Ent}_{\mathrm{Poi}(\lambda)}(f) \le \lambda\mathbb{E}_{\mathrm{Poi}(\lambda)}(A(f, \mathrm{D}f)), \tag{7.10}$$

and in particular

$$\mathrm{Ent}_{\mathrm{Poi}(\lambda)}(f) \le \lambda\mathbb{E}_{\mathrm{Poi}(\lambda)}(B(f, \mathrm{D}f)), \tag{7.11}$$

and

$$\mathrm{Ent}_{\mathrm{Poi}(\lambda)}(f) \le \lambda\mathbb{E}_{\mathrm{Poi}(\lambda)}(C(f, \mathrm{D}f)). \tag{7.12}$$

Equality is achieved in (7.11) and in (7.12) when $f(n) = \mathrm{e}^{-\alpha n}$, $n \in \mathbb{N}$, in the limit as $\alpha \to 0$. The constant λ is thus optimal, in other words minimal.

The ratio between the optimal constants in (7.10) or (7.12) and in (7.9) is 1 and not 2. This fact is related to the absence of the chain rule on $\mathbb{N}$.

Proof. It suffices to proceed as in the proof of the Poincaré inequality (7.9). □

Remark (From Poisson to Gauss laws)**.** *By using the stability by convolution of Poisson laws and the central limit theorem for i.i.d. Poisson random variables, it is possible to deduce the optimal Poincaré inequality for the Gauss law on $\mathbb{R}$ from the Poincaré inequality* (7.9) *for Poisson laws. The same method allows to deduce the optimal logarithmic Sobolev inequality for the Gauss measure, in its $\mathbb{L}^1$ form, from the modified logarithmic Sobolev inequality* (7.10) *for Poisson laws. However, the same method used with the modified logarithmic Sobolev inequality* (7.11) *or* (7.12) *leads to a logarithmic Sobolev inequality for the Gauss law on $\mathbb{R}$ with a constant equal to twice the optimal one. This is due to the fact that the comparisons* (7.4) *are not optimal, by a factor of* 2*, when $v \to 0$.*

7.2.1 Poisson process and A modified inequality

Let us show that we can prove the modified entropic inequality (7.10) using semigroup interpolation. Let $X = (X_t)_{t\geq 0}$ be a simple Poisson process with intensity $\lambda > 0$. It is of course a Lévy process on $\mathbb{R}$, but prefer to see it here as a continuous time Markov chain on $\mathbb{N}$. Let $P = (P_t)_{t\geq 0}$ be its Markov semigroup. Here X and P play for the Poisson law the role played for the Gauss law by Brownian motion and the heat semigroup. For any $t \geq 0$ and $x \in \mathbb{N}$ we have

$$\mathrm{Law}(X_t \mid X_0 = x) = \delta_x * \mathrm{Poi}(\lambda t) = \mathrm{e}^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}\delta_{x+n}$$

and $P_t(f)(x) = \mathbb{E}(f(X_t) \mid X_0 = x) = \mathbb{E}(f(x + X_t))$ with $X_t \sim \mathrm{Poi}(\lambda t)$. The infinitesimal generator of P is given, for $f : \mathbb{N} \to \mathbb{R}$ and $x \in \mathbb{N}$, by

$$(Lf)(x) = \partial_{t=0}P_t(f)(x) = \lambda((f(x + 1) - f(x)) = \lambda(\mathrm{D}f)(x).$$

We have $LP_t = P_tL$. It can be check that we have the commutation formulas

$$\mathrm{D}L = \mathrm{LD} \quad \text{and} \quad \mathrm{D}P_t(f) = P_t(\mathrm{D}f).$$

Now for any bounded $f : \mathbb{N} \to (0, \infty)$ we have, with $\phi(u) := u\log(u)$,

$$\mathrm{Ent}_{\mathrm{Poi}(\lambda t)}(f) = P_t(\phi(f)) - \phi(P_t(f)) = \int_0^t \partial_s(P_s(\phi(P_{t-s}f)))\mathrm{d}s.$$

Setting $g := P_{t-s}f$ we get, using the function A as in (7.5),

$$\partial_s(P_s(\phi(P_{t-s}f))) = P_s(L\phi(g) - \phi'(g)Lg) = \lambda P_s(A(g, \mathrm{D}g)).$$

Hence the semigroup interpolation leads to the A function, and not to B or C. Using the commutation $\mathrm{D}g = P_{t-s}(\mathrm{D}f)$, the Jensen inequality for the convex function A and the probability measure $P_{t-s}(\cdot)$, and the semigroup property,

$$P_s(A(g, \mathrm{D}g)) = P_s(A(P_{t-s}f, P_{t-s}(\mathrm{D}f))) \leq P_s(P_{t-s}(A(f, \mathrm{D}f))) = P_t(A(f, \mathrm{D}f)).$$

This gives finally the A-based modified logarithmic Sobolev inequality

$$\mathrm{Ent}_{\mathrm{Poi}(\lambda t)}(f) \leq \lambda t\mathbb{E}_{\mathrm{Poi}(\lambda t)}(A(f, \mathrm{D}f)).$$

The lack of chain rule in discrete space produces a lack of diffusion property, which is circumvented here by using convexity and the Jensen inequality.

This A-based modified logarithmic Sobolev inequality can be generalized far beyond $\mathbb{N}$ to general Poisson point processes by using suitable tools and concepts from stochastic calculus. These ideas are explored by Luming Wu in [33].

7.2.2 M/M/Infinity queue and B modified inequality

Let us show that we can prove the modified entropic inequality (7.11) using semigroup interpolation. Let $X = (X_t)_{t\geq 0}$ be the M/M/∞ queuing process with intensities $\lambda > 0$ and $\mu > 0$. It is a continuous time Markov chain on $\mathbb{N}$. Let $P = (P_t)_{t\geq 0}$ be its Markov semigroup. Set $\rho := \lambda/\mu$. Here X and P play for the Poisson law $\pi_\rho := \mathrm{Poi}(\rho)$ the role played for the Gauss law by the Ornstein–Uhlenbeck process and its semigroup. For any

$t \geq 0$ and $x \in \mathbb{N}$ we have the following discrete and Poisson-Binomial analogue of the Mehler formula

$$\mathrm{Law}(X_t \mid X_0 = x) = \mathrm{Bin}(x, \mathrm{e}^{-\mu t}) * \mathrm{Poi}(\rho(1 - \mathrm{e}^{-\mu t})).$$

In particular this shows that the Poisson law π_ρ is stationary in the sense that $\mathrm{Law}(X_t \mid X_0 = x) \to \pi_\rho$ weakly, as $t \to \infty$, for any $x \in \mathbb{N}$, and also invariant in the sense that if $X_0 \sim \mathrm{Poi}(\rho)$ then $X_t \sim \mathrm{Poi}(\rho)$ for any $t \geq 0$. The invariance can be writen as $\mathbb{E}_{\pi_\rho}(P_t f) = \mathbb{E}_{\pi_\rho}(f)$ for any bounded $f : \mathbb{N} \to \mathbb{R}$ and any $t \geq 0$.

The infinitesimal generator of P is given, for $f : \mathbb{N} \to \mathbb{R}$ and $x \in \mathbb{N}$, by

$$\begin{aligned}(Lf)(x) &= \partial_{t=0} P_t(f)(x) \\ &= \lambda((f(x+1) - f(x)) + x\mu(f(x-1) - f(x)) \\ &= \lambda(\mathrm{D}f)(x) + x\mu(\mathrm{D}^* f)(x).\end{aligned}$$

It is a continuous time birth and death process, with birth rate λ and death rate $x\mu$. We have $P_t L = L P_t$. It can be checked that we have the commutations

$$\mathrm{D}L = \mathrm{LD} - \mu \mathrm{D} \quad \text{and} \quad \mathrm{D}P_t(f) = \mathrm{e}^{-\mu t} P_t(\mathrm{D}f).$$

It can be checked that π_ρ is reversible. This means that if $X_0 \sim \pi_\rho$ then the random couples (X_0, X_t) and (X_t, X_0) have the same law for any $t \geq 0$. This gives a "Poisson" integration by parts formula: for any bounded $f, g : \mathbb{N} \to \mathbb{R}$,

$$\mathbb{E}_{\pi_\rho}(fLg) = \mathbb{E}_{\pi_\rho}(gLf) = -\lambda \mathbb{E}_{\pi_\rho}((\mathrm{D}f)(\mathrm{D}g)).$$

Following [14], for any $t \geq 0$ and bounded $f : \mathbb{N} \to (0, \infty)$, denoting $h := P_t f$ and $\phi(u) := u \log(u)$, we get, by using invariance and integration by parts,

$$-\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ent}_{\pi_\rho}(P_t f) = \mathbb{E}_{\pi_\rho}(\phi'(h) L h) = \lambda \mathbb{E}_{\pi_\rho}(\mathrm{D}(\phi'(h)) \mathrm{D}h) = \lambda \mathbb{E}_{\pi_\rho}(B(h, \mathrm{D}h)).$$

Hence the exponential decay of the entropy along the time leads to the B function, and not to A or C. Now the commutation gives $\mathrm{D}h = \mathrm{e}^{-\mu t} P_t(\mathrm{D}f)$ and thus, by using the Jensen inequality for the convex function B and $B(u, cv) \leq cB(u, v)$ when $c \in [0, 1]$[1],

$$B(h, \mathrm{D}h) = B(P_t f, \mathrm{e}^{-\mu t} P_t \mathrm{D}f) \leq \mathrm{e}^{-\mu t} P_t(B(f, \mathrm{D}f)).$$

Therefore, using the invariance of π_ρ,

$$\mathbb{E}_{\pi_\rho}(B(h, \mathrm{D}h)) \leq \mathrm{e}^{-\mu t} \mathbb{E}_{\pi_\rho}(P_t(B(f, \mathrm{D}f))) = \mathrm{e}^{-\mu t} \mathbb{E}_{\pi_\rho}(B(f, \mathrm{D}f)).$$

Since $\lim_{t \to \infty} \mathrm{Ent}_{\pi_\rho}(P_t f) = 0$, we get finally (7.11) by integrating over t, namely

$$\begin{aligned}\mathrm{Ent}_{\pi_\rho}(P_t f) &= -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ent}_{\pi_\rho}(P_t f) \, \mathrm{d}t \\ &\leq \lambda \left(\int_0^\infty \mathrm{e}^{-\mu t} \, \mathrm{d}t \right) \mathbb{E}_{\pi_\rho}(B(f, \mathrm{D}f)) \\ &= \rho \mathbb{E}_{\pi_\rho}(B(f, \mathrm{D}f)).\end{aligned}$$

The sole inequality comes from the Jensen inequality for B and P_t.

[1] Indeed, $\Phi'(u) = 1 + \log(u)$ is non-decreasing so $B(u, cv) = (\Phi'(u+cv) - \Phi'(u))cv \leq (\Phi'(u+v) - \Phi'(u))cv = cB(u, v)$ if $v \geq 0$, and a careful examination reveals that the same holds actually when $v \leq 0$.

Remark (More general Markov chains)**.** *Following [13], let $(X_t)_{t\geq 0}$ be a continuous time Markov chain with at most countable state space E, irreducible, positive recurrent, and aperiodic, with unique invariant probability measure π, and with infinitesimal generator $L : E \times E \to \mathbb{R}$. We have, for every $x, y \in E$,*

$$L(x,y) = \partial_{t=0}\mathbb{P}(X_t = y \mid X_0 = x).$$

We see L as matrix with non-negative off-diagonal elements and zero-sum rows: $L(x,y) \geq 0$ and $L(x,x) = -\sum_{y\neq x} L(x,y)$ for every $x, y \in E$. The invariance reads $0 = \sum_{x\in E} \pi(x)L(x,y)$ for every $y \in E$. The operator L acts on functions as $(Lf)(x) = \sum_{y\in E} L(x,y)f(y)$ for every $x \in E$. We have $\pi(x) > 0$ for every $x \in E$, and for any probability measure μ on E, denoting $f(x) := \mu(x)/\pi(x)$,

$$\mathrm{H}(\mu \mid \pi) = \mathrm{Ent}_\pi(f) = \sum_{x\in E} \phi\Big(\frac{\mu(x)}{\pi(x)}\Big)\pi(x).$$

We can see $x \mapsto \mu(x)$ as a density with respect to the counting measure on E. For any $t \geq 0$, if $\mu_t(x) := \mathbb{P}(X_t = x)$ then $g_t(x) := \mu_t(x)/\pi(x)$ and $\partial_t g_t = L^* g_t$ where L^* is the adjoint of L in $\ell^2(\pi)$ given by $L^*(x,y) = L(y,x)\pi(y)/\pi(x)$. Now

$$\partial_t \mathrm{H}(\mu_t \mid \pi) = \sum_{x\in E} [\phi'(g_t)L^* g_t](x)\pi(x).$$

The right hand side is up to a sign a discrete Fisher information. Moreover

$$\partial_t^2 \mathrm{H}(\mu_t \mid \pi) = \sum_{x\in E} \Big[g_t LL \log(g_t) + \frac{(L^* g_t)^2}{g_t}\Big](x)\pi(x).$$

The right hand side can be nicely rewritten when π is reversible, and constitutes a discrete analogue of a Γ_2 formula for diffusions. The lack of chain rule in discrete spaces explains the presence of two distinct terms in the right hand side.

7.2.3 Concentration and C modified inequality

Let us show that the modified entropic inequality (7.12) implies a "sub-Poissonian tailed" concentration of measure for discrete Lipschitz functions.

Theorem 7.8 (Sub-Poissonian concentration of measure)**.** *Let μ be a probability measure on $\mathbb{N}$ such that for some $c > 0$ and for any $f : \mathbb{N} \to (0,\infty)$,*

$$\mathrm{Ent}_\mu(f) \leq c\mathbb{E}_\mu\Big(\frac{(\mathrm{D}f)^2}{f}\Big)$$

where $(\mathrm{D}f)(x) = f(x+1) - f(x)$ for any $x \in \mathbb{N}$. Then for any $F : \mathbb{N} \to [0,\infty)$ such that $\sup_{x\in\mathbb{N}} |(\mathrm{D}F)(x)| \leq 1$ and $\mathbb{E}_\mu(F) < \infty$, we have, for any $r \geq 0$,

$$\mu(F \geq \mathbb{E}_\mu(F) + r) \leq \exp\Big(-\frac{r}{4}\log\Big(1 + \frac{r}{2c}\Big)\Big).$$

In particular $\mathbb{E}_\mu(\mathrm{e}^{\alpha|F|\max(0,\log|F|)}) < \infty$ for a small enough $\alpha > 0$.

Let us compare with the Chernoff bound in the case of a Poisson probability measure $\mu = \mathrm{Poi}(\lambda)$ when $F(x) = x$ for any $x \in \mathbb{N}$. In this case $\mathbb{E}_\mu(F) = c = \lambda$. For any $r > 0$, with $X \sim \mathrm{Poi}(\lambda)$, by using the Markov inequality,

$$\mu(F \geq \mathbb{E}_\mu(F) + r) = \mathbb{P}(X \geq \lambda + r)$$

$$
\begin{aligned}
&\leq \mathrm{e}^{-\inf_{\theta>0}(\theta(\lambda+r)-\log\mathbb{E}(\mathrm{e}^{\theta X}))}\\
&= \mathrm{e}^{-\inf_{\theta>0}(\theta(\lambda+r)-\lambda(\mathrm{e}^{\theta}-1))}\\
&= \mathrm{e}^{-(\lambda+r)\log(1+\frac{r}{\lambda})+r}.
\end{aligned}
$$

The $r\log(r)$ behavior in the right hand side is asymptotically optimal as $r\to\infty$.

Proof. Following Michel Ledoux [23, Prop. 5.4], we use the Herbst argument, as for the logarithmic Sobolev inequalities for the Gauss law. Let $F:\mathbb{N}\to[0,\infty)$ be such that $\sup_{x\in\mathbb{N}}|(\mathrm{D}F)(x)|\leq 1$. We would like to bound the Laplace transform

$$\lambda\geq 0\mapsto H(\lambda):=\mathbb{E}_\mu(\mathrm{e}^{\lambda F}).$$

The assumed modified logarithmic Sobolev inequality for μ gives, for $f=\mathrm{e}^{\lambda F}$,

$$\mathrm{Ent}_\mu(\mathrm{e}^{\lambda F})\leq c\mathbb{E}_\mu\Big(\frac{(\mathrm{D}(\mathrm{e}^{\lambda F}))^2}{\mathrm{e}^{\lambda F}}\Big).$$

Now we have the following ersatz of chain rule, for any $g:\mathbb{N}\to\mathbb{R}$,

$$|\mathrm{D}(\mathrm{e}^g)|\leq|\mathrm{D}g|\mathrm{e}^{|\mathrm{D}g|}\mathrm{e}^g,$$

since by the intermediate value theorem, for any $x\in\mathbb{N}$,

$$|\mathrm{D}(\mathrm{e}^g)(x)|=|\mathrm{e}^{g(x+1)}-\mathrm{e}^{g(x)}|=|\mathrm{D}g(x)|\mathrm{e}^\tau$$

for some $\tau\in(g(x)\wedge g(x+1),g(x)\vee g(x+1))$, and $\tau\leq g(x)+|(\mathrm{D}g)(x)|$. Thus,

$$\mathrm{Ent}_\mu(\mathrm{e}^{\lambda F})\leq c\lambda^2\mathrm{e}^{2\lambda}\mathbb{E}_\mu(\mathrm{e}^{\lambda F})=c\lambda^2\mathrm{e}^{2\lambda}H(\lambda).$$

Now if we define $K(\lambda):=\frac{1}{\lambda}\log H(\lambda)$ then

$$\mathrm{Ent}_\mu(\mathrm{e}^{\lambda F})=\lambda H'(\lambda)-H(\lambda)\log H(\lambda)=\lambda^2H(\lambda)K'(\lambda).$$

Thus $K'(\lambda)\leq c\mathrm{e}^{2\lambda}$. But $H(0)=1$ and $K(0)=H'(0)=\mathbb{E}_\mu(F)$ (exists!), hence

$$K(\lambda)\leq K(0)+c\frac{\mathrm{e}^{2\lambda}-1}{2},\quad\text{and}\quad H(\lambda)\leq\mathrm{e}^{\lambda\mathbb{E}_\mu(F)+\frac{c\lambda}{2}(\mathrm{e}^{2\lambda}-1)}.$$

It follows, thanks to the Markov inequality, that for any $r\geq 0$ and $\lambda\geq 0$,

$$\mu(F\geq\mathbb{E}_\mu F+r)\leq\mathrm{e}^{-\lambda r-\lambda\mathbb{E}(F)}H(\lambda)\leq\mathrm{e}^{-\lambda r+\frac{c\lambda}{2}(\mathrm{e}^{2\lambda}-1)}.$$

When $r\leq 2c$ (the constants are not sharp!), taking $\lambda=\frac{r}{4c}\leq\frac{1}{2}$ gives[2]

$$\mathrm{e}^{-\lambda r+\frac{c\lambda}{2}(\mathrm{e}^{2\lambda}-1)}\leq\mathrm{e}^{-\lambda r+2c\lambda^2}\leq\mathrm{e}^{-\frac{r^2}{8c}}.$$

When $r\geq 2c$, taking $\lambda=\frac{1}{2}\log(\frac{r}{c})$ gives

$$\mathrm{e}^{-\lambda r+\frac{c\lambda}{2}(\mathrm{e}^{2\lambda}-1)}\leq\mathrm{e}^{-\frac{r}{4}\log(\frac{r}{c})}.$$

$\square$

We refer to [23] and references therein for sharper estimates and further comments.

[2] We use $\mathrm{e}^{2u}-1\leq 2(\mathrm{e}-1)u\leq 4u$ when $0\leq u\leq 1/2$ (exponential curve below linear segment).

Bibliography

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer – *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux.

[2] D. Bakry and M. Ledoux – "Lévy-Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator", *Invent. Math.* **123** (1996), no. 2, p. 259–281.

[3] D. Bakry, I. Gentil and M. Ledoux – *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014.

[4] F. Barthe and B. Maurey – "Some remarks on isoperimetry of Gaussian type", *Ann. Inst. H. Poincaré Probab. Statist.* **36** (2000), no. 4, p. 419–434.

[5] F. Barthe – *Isoperimetric inequalities, probability measures and convex geometry*, Eur. Math. Soc., Zürich, 2005.

[6] S. Bobkov – "A functional form of the isoperimetric inequality for the Gaussian measure", *J. Funct. Anal.* **135** (1996), no. 1, p. 39–49.

[7] S. G. Bobkov – "An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space", *Ann. Probab.* **25** (1997), no. 1, p. 206–214.

[8] C. Borell – "The Ehrhard inequality", *C. R. Math. Acad. Sci. Paris* **337** (2003), no. 10, p. 663–666.

[9] Y. Brenier – "Décomposition polaire et réarrangement monotone des champs de vecteurs. (Polar decomposition and increasing rearrangement of vector fields)", *C. R. Acad. Sci., Paris, Sér. I* **305** (1987), p. 805–808 (French).

[10] L. A. Caffarelli – "Monotonicity properties of optimal transportation and the FKG and related inequalities", *Comm. Math. Phys.* **214** (2000), no. 3, p. 547–563.

[11] — , "Erratum: "Monotonicity of optimal transportation and the FKG and related inequalities" [Comm. Math. Phys. **214** (2000), no. 3, 547–563; MR1800860 (2002c:60029)]", *Comm. Math. Phys.* **225** (2002), no. 2, p. 449–450.

[12] M. Capitaine, E. P. Hsu and M. Ledoux – "Martingale representation and a simple proof of logarithmic Sobolev inequalities on path spaces", *Electron. Comm. Probab.* **2** (1997), p. 71–81 (electronic).

[13] P. Caputo, P. Dai Pra and G. Posta – "Convex entropy decay via the Bochner-Bakry-Emery approach", *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009), no. 3, p. 734–753.

[14] D. Chafaï – "Binomial-Poisson entropic inequalities and the $M/M/\infty$ queue", *ESAIM Probab. Stat.* **10** (2006), p. 317–339 (electronic).

[15] Y. Chen – "An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture", *Geom. Funct. Anal.* **31** (2021), no. 1, p. 34–61 (English).

[16] D. Cordero-Erausquin – "Some applications of mass transport to Gaussian-type inequalities", *Arch. Ration. Mech. Anal.* **161** (2002), no. 3, p. 257–269 (English).

[17] A. Ehrhard – "Symétrisation dans l'espace de Gauss", *Math. Scand.* **53** (1983), p. 281–301.

[18] L. Erdős and H.-T. Yau – "Dynamical approach to random matrix theory", 2017.

[19] L. Gross – "Logarithmic Sobolev inequalities", *Amer. J. Math.* **97** (1975), no. 4, p. 1061–1083.

[20] L. Gross and O. Rothaus – "Herbst inequalities for supercontractive semigroups", *J. Math. Kyoto Univ.* **38** (1998), no. 2, p. 295–318.

[21] A. Guionnet and B. Zegarlinski – "Lectures on logarithmic Sobolev inequalities", Séminaire de Probabilités, XXXVI, Lecture Notes in Math., vol. 1801, Springer, Berlin, 2003, p. 1–134.

[22] M. Ledoux – "Isoperimetry and Gaussian analysis", Lectures on probability theory and statistics (Saint-Flour, 1994), Lecture Notes in Math., vol. 1648, Springer, Berlin, 1996, p. 165–294.

[23] — , "Concentration of measure and logarithmic Sobolev inequalities", Séminaire de Probabilités, XXXIII, Lecture Notes in Math., vol. 1709, Springer, Berlin, 1999, p. 120–216.

[24] — , "The geometry of Markov diffusion generators", *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000), no. 2, p. 305–366, Probability theory.

[25] R. J. McCann – "Existence and uniqueness of monotone measure-preserving maps", *Duke Math. J.* **80** (1995), no. 2, p. 309–323 (English).

[26] R. Montenegro and P. Tetali – *Mathematical aspects of mixing times in markov chains*, Foundations and Trends in Theoretical Computer Science Series, Now Publishers, 2006.

[27] G. Royer – *Une initiation aux inégalités de Sobolev logarithmiques*, Cours Spécialisés [Specialized Courses], vol. 5, Société Mathématique de France, Paris, 1999.

[28] — , *An initiation to logarithmic Sobolev inequalities*, SMF/AMS Texts and Monographs, vol. 14, American Mathematical Society, Providence, RI; Société Mathématique de France, Paris, 2007, Translated from the 1999 French original by Donald Babbitt.

[29] L. Saloff-Coste – *Aspects of Sobolev-type inequalities*, London Mathematical Society Lecture Note Series, vol. 289, Cambridge University Press, Cambridge, 2002.

[30] F. Santambrogio – *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*, vol. 87, Cham: Birkhäuser/Springer, 2015 (English).

[31] M. Talagrand – "Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem", *Geom. and Funct. Anal.* **3** (1993), p. 295–314.

[32] C. Villani – *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new.

[33] L. Wu – "A new modified logarithmic Sobolev inequality for Poisson point processes and several applications", *Probab. Theory Related Fields* **118** (2000), no. 3, p. 427–438.