LOGARITHMIC SOBOLEV INEQUALITIES AND RELATED TOPICS

THÉO DUMONT[†]

ABSTRACT. This is a short note whose goal is to introduce logarithmic Sobolev inequalities for probability measures. Starting with Sobolev embeddings and the Sobolev inequality for the Lebesgue measure, we establish the logarithmic Sobolev inequality that adapts to the infinite-dimensional setting, and state the linear convergence properties it implies on the flow of the relative entropy. We mention some more inequalities for fun and link these inequalities to their Euclidean counterparts via the Wasserstein–Otto geometry of probability measures.
This is a learning document and some mistakes or inaccuracies are probably hidden in several places. If you spot any, feel free to signal them at theo.dumont@univ-eiffel.fr!

TOPICS: Poincaré inequality, Sobolev inequality, logarithmic Sobolev inequality, HWI inequality, Bakry–Émery condition, Holley–Stroock perturbation lemma, convergence rates for the entropy flow, ~~Markov processes~~.

CONTENTS

1. INTRODUCTION

1.1. **Convergence of gradient flows in Euclidean spaces.** Consider the Euclidean space $\mathbb{R}^n$ and flows of functions $f : \mathbb{R}^n \to \mathbb{R}$ on it, aiming to find an element

$$x^\star \in \arg\min_{x\in\mathbb{R}^n} f(x),$$

which we assume is non-empty, via

$$\dot{x}_t = -\nabla f(x_t).$$

[†]Laboratoire d'Informatique Gaspard Monge, Université Gustave Eiffel, CNRS, F-77454 Marne-la-Vallée, France.
E-mail address: theo.dumont@univ-eiffel.fr.
Date: Last revised on January 6, 2025.

Can we show that our flow will converge to a minimizing element $x^\star \in \arg\min_x f(x)$? If f is *strongly convex*, that is, $\nabla^2 f \geq KI$, or

$$f(x) - f(y) \geq \langle \nabla f(y), x - y\rangle + \frac{K}{2}\|y - x\|^2, \tag{SC}$$

then $x^\star$ is unique and

$$\frac{\mathrm{d}}{\mathrm{d}t}\|x_t - x^\star\|^2 = 2\langle x_t - x^\star, \dot{x}_t\rangle = -2\langle x_t - x^\star, \nabla f(x_t)\rangle \leq -K\|x_t - x^\star\|^2,$$

and Grönwall's lemma gives linear convergence $\|x_t - x^\star\|^2 \leq \|x_0 - x^\star\|^2 e^{-Kt}$. Actually, one merely needs the *Polyak–Łojasiewicz condition*

$$f(x) - f^\star \leq \frac{1}{2K}\|\nabla f(x)\|^2, \tag{PŁ}$$

where $f^\star = \min_x f(x)$. This implies in particular that every critical point of f is a global minimizer of it. This condition was originally introduced by Polyak [Pol63] in 1963, who showed that it is a sufficient condition for gradient descent to achieve a linear convergence rate. It is a special case of an inequality introduced in the same year by Łojasiewicz [Łoj63] (see Remark 1.5). Under this condition,

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(f(x_t) - f^\star\big) = \langle \nabla f(x_t), \dot{x}_t\rangle = -\|\nabla f(x_t)\|^2 \leq -2K\big(f(x_t) - f^\star\big),$$

and the linear convergence follows by Grönwall's lemma. Actually, linear convergence of $f(x_t)$ toward $f^\star$ for any initial point x_0 implies (PŁ), so it is an equivalence.

Proposition 1.1. *(SC)$_K$ implies (PŁ)$_K$.*

Proof. Just apply Young's inequality to the strong convexity inequality. For any $\kappa > 0$,

$$f(x) - f^\star \leq \langle \nabla f(x), x - x^\star\rangle - \frac{K}{2}\|x - x^\star\|^2 \leq \frac{1}{2\kappa}\|\nabla f(x)\|^2 + \frac{\kappa - K}{2}\|x - x^\star\|^2.$$

Note that this means that (PŁ)$_\kappa$ is always satisfied, with some small second-order error term. In particular, this is (PŁ)$_K$ when $\kappa = K$. □

Finally, let us say that f has *quadratic growth* if

$$d(x, \arg\min f)^2 \leq \frac{2}{K}\big(f(x) - f^\star\big). \tag{QG}$$

Quite easily we have the following:

Proposition 1.2. *(SC)$_K$ implies (QG)$_K$.*

But we actually have the stronger result:

Proposition 1.3. *(PŁ)$_K$ implies (QG)$_K$.*

I couldn't find a proof for this in the literature, although the result is mentioned at least by Dello Schiavo, Maas, and Pedrotti [DMP24] and Garrigos and Gower [GG23]. In [LDZ24], an additional assumption is taken ($f + \frac{\lambda}{2}\|\cdot\|^2$ should be convex for some $\lambda \in \mathbb{R}$), and in [KNS16] a stronger one (f should be L-smooth, see (Smoothness)) but it seems we don't need this. The following proof is taken from Stromme's talk [Str24] and personal communication.

Proof. Suppose that f satisfies (PŁ)$_K$. Consider the flow $\dot{x}_t = -\nabla f(x_t)$ and the Lyapunov function

$$h(t) = \sqrt{f(x_t) - f^\star} + \lambda\|x_t - x_0\|,$$

where λ is some constant that will be chosen later to make things work. Suppose h non-increasing. Then the term $\|x_t - x_0\|$ ensures that x_t cannot escape and remains in some compact set. Up to taking a subsequence, x_t converges to some x_∞ and since $f(x_t) \to f^\star$ by (PŁ), then $x_\infty \in \arg\min f$. Then we have

$$\sqrt{f(x_0) - f^\star} = h(0) \geq \lim_{t\to\infty} h(t) = \lambda\|x_\infty - x_0\| \geq \lambda d(x_0, \arg\min f),$$

where we used the decrease of h and the fact that $x_\infty \in \arg\min f$. This is (QG) with constant $2\lambda^2$. To find a λ such that h is non-increasing, compute

$$\begin{aligned} h'(t) &= \frac{-\|\nabla f(x_t)\|^2}{2\sqrt{f(x_t)-f^\star}} - \lambda \frac{\langle \nabla f(x_t), x_t - x_0\rangle}{\|x_t - x_0\|} \\ &\le \frac{-\sqrt{2K}\sqrt{f(x_t)-f^\star}\|\nabla f(x_t)\|}{2\sqrt{f(x_t)-f^\star}} + \lambda\|\nabla f(x_t)\| \quad \text{by (PŁ) and Cauchy–Schwarz's inequality} \\ &= \big(\lambda - \sqrt{K/2}\big)\|\nabla f(x_t)\|. \end{aligned}$$

Taking $\lambda = \sqrt{K/2}$ yields $h'(t) \le 0$, hence f satisfies (QG) with constant $2(\sqrt{K/2})^2 = K$. ☐

So far, we have:

$$\begin{array}{ccccc} \|\cdot\| \lesssim e^{-Kt} & & f(\cdot) - f^\star \lesssim e^{-Kt} & & \\ \Uparrow & & \Updownarrow & & \\ (\text{SC})_K & \Longrightarrow & (\text{PŁ})_K & \overset{(*)}{\Longrightarrow} & (\text{QG})_K \end{array}$$

where $(*)$ is an equivalence if f is convex. Actually, if a function f satisfies the quadratic growth condition (QG), linear convergence of the values of f trivially implies linear convergence of the x's. Since (PŁ) implies (QG), then (PŁ) implies linear convergence of the x's as well. See [KNS16; LDZ24] for more details on these conditions and some other ones as well, and the very nice and detailed [GG23].

Remark 1.4 (Stability of (PŁ) under bounded change of geometry [Str24]) If we endow $\mathbb{R}^n$ with a metric tensor g that satisfies $\tilde{K}I \le g$, then $g^{-1} \le \tilde{K}^{-1}I$ and

$$\|\operatorname{grad}_g f\|^2 = \langle g^{-1}\nabla f, \nabla f\rangle \le \frac{1}{\tilde{K}}\|\nabla f\|^2.$$

This means that if f satisfies $(\text{PŁ})_{K/\tilde{K}}$ in the standard geometry, then it satisfies $(\text{PŁ})_K$ in the geometry induced by g. This is also what we do in Theorem 2.6 for the (Brascamp–Lieb) and (Poincaré) inequalities. △

Remark 1.5 (Generalizations of (PŁ)) The (PŁ) inequality is a special case of the Łojasiewicz [Łoj63] inequality

$$\big(f(x) - f^\star\big)^\theta \le \frac{1}{\sqrt{2K}}\|\nabla f(x)\|, \tag{Ł}$$

where $\theta \in [\frac{1}{2}, 1)$. (PŁ) corresponds to choosing $\theta = \frac{1}{2}$ in (Ł). In the special case where f is convex (hence weakly-convex and (PŁ) implies convergence of the x's), then (Ł) yields convergence in $O(t^{-\frac{1-\theta}{2\theta-1}})$ when $\frac{1}{2} < \theta < 1$. Assume $f^\star = 0$. Then the (Ł) inequality itself is a special case of the Kurdyka–Łojasiewicz [Kur98] inequality

$$1 \le \|\nabla(\varphi \circ f)(x)\|, \tag{KŁ}$$

where φ is a *desingularizing function*: C^1, concave, and such that $\varphi(0) = 0$ and $\varphi' > 0$. In other words, up to a reparametrization, f is sharp around $x^\star$. (Ł) corresponds to choosing $\varphi(t) = \frac{\sqrt{2K}}{(1-\theta)}t^{1-\theta}$ in (KŁ). Regarding the convergence rate: let Φ be such that $\Phi' = -\varphi'^2$. Convergence of the $f(x)$'s is in $O(\Phi^{-1}(t - t_1))$ and convergence of the x's is in $O(\varphi \circ \Phi^{-1}(t - t_1))$ for some $t_1 \in \mathbb{R}$ [Gar15, Theorem 3.1.12].

See also [FG21] for a relaxation of (PŁ) of the form

$$g(x) - \xi \le \frac{1}{2K}\|\nabla f(x)\|^\alpha, \tag{proxy-PŁ}$$

which allows for the existence of stationary points that are not globally minimizing via the use of a proxy function g, [LZB22] for a study of (PŁ) being satisfied merely on a subset of $\mathbb{R}^n$, or [BPV22] for a local (PŁ) inequality guaranteeing convergence given good initial conditions. See also [DMP24] for more information. △

Remark 1.6 (Lower bound on the convergence rate for convex functions) If we assume f to be merely convex (and not strongly-convex for instance), one cannot expect to get a better convergence rate than $O(\frac{1}{t^2})$, see for instance [Bub+15, Theorem 3.14]. △

Remark 1.7 (Convergence of gradient descent) In practice, we implement the *gradient descent*

$$x_{t+1} = x_t - \eta_t \nabla f(x_t),$$

where t takes integer values and η_t is the step size. To prove convergence of the values of f to the minimum value in this discrete scheme, in addition to the (PL) condition one must also assume *smoothness* of f, that is, $\nabla^2 f \leq LI$, or

$$f(x) - f(y) \leq \langle \nabla f(y), y - x \rangle + \frac{L}{2}\|x - y\|^2 \qquad \text{(Smoothness)}$$

(compare with (SC), which is the opposite bound on the Hessian). If this is satisfied, then

$$\begin{aligned} f(x_{t+1}) - f^\star &= f(x_t - \eta \nabla f(x_t)) - f^\star \\ &\leq f(x_t) - \langle \nabla f(x_t), \eta \nabla f(x_t) \rangle + \frac{L}{2}\|\eta \nabla f(x_t)\|^2 - f^\star = f(x_t) - f^\star + \eta\Big(\frac{L\eta}{2} - 1\Big)\|\nabla f(x_t)\|^2 \end{aligned}$$

Now, if $0 \leq \eta \leq 2/L$, then $\frac{L\eta}{2} - 1 \leq 0$ and one can apply (PŁ) to get

$$f(x_{t+1}) - f^\star \leq \big(1 - K\eta(2 - L\eta)\big)\big(f(x_t) - f^\star\big) = (1 - K/L)\big(f(x_t) - f^\star\big)$$

when choosing the optimal value $\eta = \frac{1}{L}$. Hence

$$f(x_t) - f^\star \leq (1 - K/L)^t\big(f(x_0) - f^\star\big) \leq e^{-\frac{K}{L}t}\big(f(x_0) - f^\star\big).$$

1.2. **Convergence of the entropy flow in $\mathcal{P}(\mathbb{R}^n)$.** [MV00] In this note, we're interested in these kinds of convergence rates but in the space $\mathcal{P}(\mathbb{R}^n)$ of (smooth) probability measures, for the flow of the entropy functional $H(\rho \,|\, e^{-V}) = \int \rho \log(\rho/e^{-V})$, which is

$$\partial_t \rho_t = \operatorname{div}(\rho_t \nabla(V + \log \rho_t)) = \operatorname{div}(\rho_t \nabla V + \nabla \rho_t). \qquad \text{(Fokker–Planck)}$$

Question 1.8. Take ρ_t solution of (Fokker–Planck) with target measure $\gamma = e^{-V}$. Do we have a linear convergence rate to γ in terms of some distance/divergence that is still to be chosen?

The stationary state is $\gamma = e^{-V}$ (up to adding a constant to V, assume that e^{-V} is a probability distribution). Change variables by writing $\rho = he^{-V}$, i.e. setting $h = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma}$, and the equation on h is:

$$\partial_t h = \Delta h - \nabla V \cdot \nabla h \qquad \text{(Fokker–Planck*)}$$

To quantify the convergence of ρ towards e^{-V}, one could rather check the convergence of h towards 1, and, for instance, check the evolution of some L^2-norm of $h - 1$. By (Fokker–Planck*),

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^n} (h-1)^2 e^{-V} = -2\int_{\mathbb{R}^n} \|\nabla h\|^2 e^{-V},$$

so maybe the $L^2(e^{-V})$-norm is the one to look at? If we had some kind of inequality like

$$\int_{\mathbb{R}^n} (h-1)^2 e^{-V} \leq C\int_{\mathbb{R}^n} \|\nabla h\|^2 e^{-V},$$

we could use Grönwall's lemma to get

$$h_0 \in L^2(e^{-V}) \implies \|h_t - 1\|_{L^2(e^{-V})} \leq e^{-\frac{1}{C}t}\|h_0 - 1\|_{L^2(e^{-V})},$$

or equivalently

$$\rho_0 \in L^2(e^{V}) \implies \|\rho_t - 1\|_{L^2(e^{V})} \leq e^{-\frac{1}{C}t}\|\rho_0 - 1\|_{L^2(e^{V})}.$$

But while the functional space $L^2(e^{-V})$ is natural at the level of (Fokker–Planck*), $L^2(e^V)$ at the level of (Fokker–Planck) is not. We could instead consider a variety of functionals controlling the distance between h and 1. Intuitively, we'd like to stay as close as possible to the space $L^1(\mathrm{d}x)$, that is, assuming finite mass / using the TV norm. For instance, we could replace $\phi(h) = (h-1)^2$ by $\phi(h) = h \log h - h + 1$, which yields the relative entropy

$$\int_{\mathbb{R}^n} \phi(h) e^{-V} = \int_{\mathbb{R}^n} \rho(\log \rho + V) =: H(\rho \,|\, e^{-V})$$

This is a nice candidate for controlling the distance between two probability distributions because of the *Csiszár–Kullback–Pinsker* inequality [Csi63; Kul59; Pin64]

$$\|\rho - \gamma\|_{\mathrm{TV}}^2 \leq \frac{1}{2} H(\rho \,|\, \gamma). \tag{Pinsker}$$

1.3. **Disclaimer and references.** This note borrows shamelessly from slides by Ledoux [Led00], lecture notes by Rigollet [Rig22], the book of Bakry, Gentil, and Ledoux [BGL14], the nice review paper by Markowich and Villani [MV00], and other references cited below. See also Villani et al.'s monographs, [Vil03, Chapter 9] and [Vil+09, Chapter 21]

This note does not aim to be an exhaustive review of logarithmic Sobolev inequalities and related, as it is merely a learning document. See the above references for more details.

2. From functional inequalities to entropy-information inequalities

2.1. Poincaré inequality.

Proposition 2.1 (p-Poincaré inequality)**.** *Let $1 \leq p < \infty$ and $\Omega \subseteq \mathbb{R}^n$ bounded connected open with Lipschitz boundary. Then there exists $C_{\Omega,p} > 0$ such that for every $f \in W^{1,p}(\Omega)$,*

$$\|f - \bar{f}\|_{L^p(\Omega)} \leq C_{\Omega,p} \|\nabla f\|_{L^p(\Omega)},$$

where $\bar{f} = \frac{1}{|\Omega|} \int_\Omega f(x) \,\mathrm{d}x$ is the average value of f over Ω.

In the case $p = 2$, $\|f - \bar{f}\|_{L^2(\Omega)}^2 = \int f^2 \,\mathrm{d}x - \big(\int f \,\mathrm{d}x\big)^2$, which is the variance of f with respect to the Lebesgue measure. More generally, write

$$\mathrm{Var}_\gamma(f) := \int_{\mathbb{R}^n} f^2 \,\mathrm{d}\gamma - \Big(\int_{\mathbb{R}^n} f \,\mathrm{d}\gamma\Big)^2$$

and say that a measure γ satisfies a *Poincaré inequality* (or *spectral gap*) with constant $K > 0$ if for all functions f,

$$\mathrm{Var}_\gamma(f) \leq \frac{1}{K} \int_{\mathbb{R}^n} \|\nabla f\|^2 \,\mathrm{d}\gamma. \tag{2.1}$$

Remark 2.2 (Why spectral gap?) Take f an eigenfunction of $-\Delta$ with positive eigenvalue $\lambda > 0$. Then applying (Poincaré):

$$\int_{\mathbb{R}^n} f^2 \,\mathrm{d}x - 0 \leq \frac{1}{K} \int_{\mathbb{R}^n} f(-\Delta f) \,\mathrm{d}x = \frac{\lambda}{K} \int_{\mathbb{R}^n} f^2 \,\mathrm{d}x,$$

hence $\lambda \geq K$ and the spectrum of the symmetric positive operator $-\Delta$ is included in $\{0\} \cup [K, \infty)$: (Poincaré) describes a gap in the spectrum of Δ. The same reasoning can be done with a general measure $\gamma = e^{-V}$ instead of the Lebesgue measure, replacing the operator Δ by $L = \Delta + \nabla V \cdot \nabla$. △

Remark 2.3 (Spectral gap on manifolds) (Warning: approximate) As such, a whole body of work focuses on estimating the first positive eigenvalue of (minus) the (Hodge) Laplacian Δ for functions (0-forms) on general manifolds M, with constants that depend on the curvature characteristics of the manifold. It is known that there exists a constant $C(n, D, \kappa) > 0$ that bounds this eigenvalue below, where n, D and κ are the dimension, upper bound on the diameter and lower bound on the Ricci curvature of M, respectively.

The Hodge Laplacian $\Delta = \delta d + d\delta$ is defined more generally on k-forms, and one could write a Poincaré inequality for k-forms:

$$\inf_{\eta} \int_M \|\omega - \eta\|^2 \,\mathrm{vol} \leq \frac{1}{K} \int_M \left(\|\delta\omega\|^2 + \|d\omega\|^2\right) \mathrm{vol} \tag{2.2}$$

for a k-form $\omega \in \Omega^k(M)$, where the infimum in the left-hand side runs over all harmonic k-forms η on M. See [GM73] for the case $k \geq 1$ and [HM24] for the case $k = 1$ in dimension 4. Note that in the case $k = 0$, (2.2) reduces to (2.1) since the minimizer of the left-hand side is $\eta^\star = (\int f \,\mathrm{vol})^2$. See also [BFP16; BFP22]. △

To express this condition with probability measures, just take $f = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma}$ to get

$$\mathrm{Var}_\gamma\left(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right) \leq \frac{1}{K} \int_{\mathbb{R}^n} \left\|\nabla \frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right\|^2 \mathrm{d}\gamma,$$

which is the Poincaré inequality with constant $K > 0$ for probability measures

$$\chi^2(\rho \,|\, \gamma) \leq \frac{1}{K} \mathcal{E}_\gamma(\rho) \tag{Poincaré}$$

for all probability measures ρ, where we set $\mathcal{E}_\gamma(\rho) \coloneqq \int \|\nabla \frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\|^2 \,\mathrm{d}\gamma$ and used

$$\mathrm{Var}_\gamma\left(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right) = \int_{\mathbb{R}^n} \left(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right)^2 \mathrm{d}\gamma - \left(\int_{\mathbb{R}^n} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma} \,\mathrm{d}\gamma\right)^2 = \int_{\mathbb{R}^n} \left(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right)^2 \mathrm{d}\gamma - 1 \eqqcolon \chi^2(\rho \,|\, \gamma).$$

It can be shown that the standard Gaussian measure γ_n satisfies (Poincaré) with constant $K = 1$.

Since a direct computation yields $\frac{\mathrm{d}}{\mathrm{d}t}\chi^2(\rho \,|\, \gamma) = -\mathcal{E}_\gamma(\rho)$ along the flow of (Fokker–Planck), Grönwall's lemma yields a first answer to our Question 1.8 at the beginning:

Proposition 2.4 (Convergence in χ^2 under Poincaré)**.** *If γ satisfies (Poincaré)$_K$, then for ρ_t solution of (Fokker–Planck)*

$$\chi^2(\rho_t \,|\, \gamma) \leq \chi^2(\rho_0 \,|\, \gamma) e^{-t/2K}.$$

Actually, the converse also holds [Van14].

Remark 2.5 (Reformulation of (Poincaré)) The (Poincaré) inequality can be rewritten as *modified logarithmic Sobolev inequalities*, which are usual logarithmic Sobolev inequalities with a log-Lipschitz constraint on the test measures. More precisely, the (Poincaré) inequality is equivalent to the existence of $c, K \geq 0$ such that for any Lipschitz probability measure ρ,

$$\|\nabla \log \rho\| \geq c \implies H(\rho \,|\, \gamma) \leq \frac{1}{K} I(\rho \,|\, \gamma). \tag{log-Sobolev (log-Lip)}$$

See [Vil+09, Theorem 22.25] and [BL97]. △

Consider $\gamma = e^{-V}$ where V is strongly convex, that is, $(\nabla^2 V)^{-1}$ exists. We say that γ satisfies a *Brascamp–Lieb inequality* with constant $K > 0$ if for all f,

$$\mathrm{Var}_\gamma(f) \leq K \int_{\mathbb{R}^n} \langle (\nabla^2 V)^{-1} \nabla f, \nabla f \rangle \,\mathrm{d}\gamma. \tag{Brascamp–Lieb}$$

Note that the right-hand side is just the gradient of f in the geometry induced by the metric tensor $\nabla^2 V$; see Remark 1.4.

Theorem 2.6. $\big[$(Brascamp–Lieb)$_K$ *and* $\nabla^2 V \geq \tilde{K} I\big]$ *implies* (Poincaré)$_{K/\tilde{K}}$.

Proof. Direct since $\langle (\nabla^2 V)^{-1} \nabla f, \nabla f \rangle \leq \frac{1}{\tilde{K}} \|\nabla f\|^2$. □

Remark 2.7 (Sharp constants)**.** In all theorems of this note that states that one inequality implies another one, we don't really care about sharp constants. For instance, it could be the case that the (Poincaré) inequality with constant $K/\tilde{K}$ satisfied above is also satisfied for some $\kappa < K/\tilde{K}$. △

Remark 2.8 (Mirror Langevin and Brascamp–Lieb) The Poincaré inequality can also be stated under a more general form

$$\mathrm{Var}_\gamma(f) \le K\mathcal{E}_\gamma(f),$$

where $\mathcal{E}_\gamma$ is the Dirichlet energy associated with some Markov process. For the infinitesimal generator $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ associated to the Langevin diffusion

$$\mathrm{d}X_t = -\nabla V(X_t) + \sqrt{2}\,\mathrm{d}B_t,$$

the Dirichlet energy is simply $\mathcal{E}_\gamma(f) = \int \|\nabla f\|^2 \,\mathrm{d}\gamma$. For the *mirror Langevin process*, defined as

$$X_t = \nabla\varphi^*(Y_t), \qquad \mathrm{d}Y_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}[\nabla^2\varphi(X_t)]^{1/2}\,\mathrm{d}B_t,$$

the Dirichlet energy is $\mathcal{E}_\gamma(f) = \int \langle (\nabla^2\varphi)^{-1}\nabla f, \nabla f\rangle \,\mathrm{d}\gamma$ and the corresponding (general) Poincaré inequality is the *mirror Poincaré inequality*

$$\mathrm{Var}_\gamma(f) \le K \int_{\mathbb{R}^n} \langle (\nabla^2\varphi)^{-1}\nabla f, \nabla f\rangle \,\mathrm{d}\gamma. \qquad \text{(mirror Poincaré)}$$

Note that the mirror Langevin process is basically performing a mirror gradient flow $\dot{x} = -(\nabla^2\varphi)^{-1}\nabla f(x)$. When $\varphi = V$, we recover the standard Newton's method, or the *Newton–Langevin process* [Che+20], and the (general) Poincaré inequality is the (Brascamp–Lieb) inequality. △

2.2. **Sobolev inequality.**

Proposition 2.9 (Gagliardo–Nirenberg–Sobolev inequality)**.** *Let $1 \le p < n$. Then there exists $C_{n,p} > 0$ such that*

$$\|f\|_{L^{p^*}} \le C_{n,p}\|\nabla f\|_{L^p}, \qquad \text{(GNS)}$$

where $\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}$, i.e. $p^* = p\frac{n}{n-p} > p$ is the Sobolev conjugate of p. This is the same as saying that

$$\dot{W}^{1,p}(\mathbb{R}^n) \hookrightarrow L^{p^*}.[1] \qquad \text{(GNS emb.)}$$

Remark 2.10 (Why p^* for the exponent?)**.** The exponent $q = p^*$ is the only one for which the (GNS) is invariant under dilation. Indeed, suppose that the inequality $\|f\|_{L^q} \le C_{n,p}\|\nabla f\|_{L^p}$ holds for any smooth compactly supported f, where q is arbitrary. Then define for $\lambda > 0$ the rescaled function $f_\lambda(x) = f(\lambda x)$. Applying the inequality to f_λ and changing variables yields

$$\|f\|_{L^q} \le C_{n,p}\lambda^{1-\frac{n}{p}+\frac{n}{q}}\|\nabla f\|_{L^p}$$

for any such f. Then, if $1 - \frac{n}{p} + \frac{n}{q} \neq 0$, that is $q \neq p^*$, we can either take $\lambda \to 0$ or $\lambda \to \infty$ depending on the sign of $1 - \frac{n}{p} + \frac{n}{q}$ to get $\|f\|_{L^q} = 0$, which is a contradiction (see e.g. [BGL14, Section 6.1]). Dilations play a central role in the analysis of Sobolev inequalities in Euclidean space [BGL14, Section 6.1], but I don't really know why yet. △

Proposition 2.11 (Sobolev embedding [Sob38])**.**

(i) Let $k \ge 0$ be an integer and $1 \le p < \infty$ a real number. For all real numbers ℓ, q such that $k > \ell$, $p < n$ and $1 \le p < q < \infty$ and $\frac{1}{p} - \frac{k}{n} = \frac{1}{q} - \frac{\ell}{n}$, then

$$W^{k,p}(\mathbb{R}^n) \hookrightarrow W^{\ell,q}(\mathbb{R}^n), \qquad \text{(Sob. emb.)}$$

In the special case $k = 1$ and $\ell = 0$, this means that

$$W^{1,p}(\mathbb{R}^n) \hookrightarrow L^{p^*}(\mathbb{R}^n), \qquad \text{(Sob. emb.')}$$

where $p^* = p\frac{n}{n-p} > p$ is the Sobolev conjugate of p. Intuitively, if $f \in L^p(\mathbb{R}^n)$ has derivative in $L^p(\mathbb{R}^n)$, then f itself has improved local behavior, meaning that it belongs to the space $L^{p^*}(\mathbb{R}^n)$ where $p^* > p$.

[1] We say that A embeds continuously into B, written $A \hookrightarrow B$, if $A \subseteq B$ and the identity mapping $\imath : A \to B$ is continuous/bounded, i.e. there exists a constant $C > 0$ such that $\|\cdot\|_B \le C\|\cdot\|_A$.

(ii) *(Hölder version) If $n < pk$ and $\frac{1}{p} - \frac{k}{n} = -\frac{r+\alpha}{n}$, then*

$$W^{k,p}(\mathbb{R}^n) \hookrightarrow C^{r,\alpha}(\mathbb{R}^n).$$

Intuitively, the existence of sufficiently many weak derivatives implies some continuity of the classical derivatives. In the special case $r = 0$, this means that if f has its k^{th} derivative in L^p and $pk > n$, then f is continuous.

Proof.

(i) The special case $k = 1$ and $\ell = 0$ is implied by the Gagliardo–Nirenberg–Sobolev inequality (GNS). Indeed, it is enough to add the term $C_{n,p}\|f\|_{L^p}$ to the right-hand side to obtain $W^{1,p}(\mathbb{R}^n) \hookrightarrow L^{p^*}(\mathbb{R}^n)$. We can then iterate to obtain higher orders.

(ii) The special case $k = 1$ and $r = 0$ is exactly Morrey's inequality

$$\|f\|_{C^{0,\alpha}} \leq C_{n,p,\alpha}\|f\|_{W^{1,p}}. \tag{Morrey}$$

It is enough to prove it, we can then iterate to obtain higher orders. □

Remark 2.12 (More general domains) Sobolev embeddings hold for Sobolev spaces $W^{k,p}(M)$ for more general domains M, in particular on compact Riemannian manifolds, or complete Riemannian manifolds with positive injectivity radius and bounded subsectional curvature. △

Remark 2.13 (Sobolev and isoperimetric inequalities) [Dot] The L^1 (GNS emb.) inequality

$$W^{1,1}(\mathbb{R}^n) \hookrightarrow L^{n/(n-1)}(\mathbb{R}^n)$$

is equivalent to the isoperimetric inequality

$$\mathrm{Vol}(\Omega)^{\frac{n-1}{n}} \leq C\,\mathrm{Area}(\partial\Omega),$$

where the area is in reference to the volume form on $\partial\Omega$. There are also generalizations for manifolds.

Proof. To obtain the isoperimetric inequality, consider piecewise linear bump functions f_ε that approximate Ω and take $\varepsilon \to 0$. Intuitively, $\|f_\varepsilon\|$ approximates $\mathrm{Vol}(\Omega)$ and $\|\nabla f_\varepsilon\|$ the area by Stokes' formula. The converse is more involved but should be doable. □

Remark 2.14 (Infinite-dimensional spaces)**.** The special case (Sob. emb.') of the Sobolev embedding theorems states that if $f \in L^p(\mathbb{R}^n)$ has one derivative in $\in L^p(\mathbb{R}^n)$, then f itself is in $\in L^{p^*}(\mathbb{R}^n)$ where $p^* = p\frac{n}{n-p} > p$. When $n \to \infty$, $p^* \to p$ and the improvement in the local behavior of f from having a derivative in $L^p(\mathbb{R}^n)$ becomes zero. In particular, for functions defined on infinite-dimensional spaces, (Sob. emb.') doesn't give anything.

Also, it is pointless to look for an equivalent of the Lebesgue measure on infinite-dimensional spaces, as *any translation-invariant Borel measure on an infinite-dimensional separable Banach space is trivial (either infinite for all sets or zero for all sets)* (proof). Gaussian measures however do admit some nice generalizations to infinite-dimensional separable Banach spaces E. Say that a Borel measure γ is Gaussian if $L_*\gamma$ is a Gaussian on $\mathbb{R}$ for every linear functional $L \in E^*$. A way of constructing such Gaussian measures is the *abstract Wiener space*, and it is the only one by the structure theorem. How Gaussian measures change under translations is then dictated by the Cameron–Martin theorem.

So, we're looking for a variation of (GNS) that preserves information when the dimension increases, and that can adapt to Gaussian measures [Gro75]. △

2.3. **Log-Sobolev inequality.**

2.3.1. *From Sobolev to log-Sobolev.* Let us focus on a specific case of (GNS emb.), which is

$$\dot{W}^{1,2}(\mathbb{R}^n) \hookrightarrow L^q(\mathbb{R}^n), \tag{GNS emb.$_2$}$$

where $q = 2^* = \frac{2n}{n-2} > 2$. Said differently, there exists $C_n > 0$ such that for any function $f : \mathbb{R}^n \to \mathbb{R}$

$$\|f\|_{L^q}^2 = \Big(\int_{\mathbb{R}^n} |f|^q \,\mathrm{d}x \Big)^{2/q} \le C_n \int_{\mathbb{R}^n} \|\nabla f\|^2 \,\mathrm{d}x = C_n \|\nabla f\|_{L^2}^2.$$

Recall that in the limit $n \to \infty$, we don't get anything from (GNS emb.$_2$) since $q \to 2$. The sharp constant is $C_n = \frac{1}{\pi n(n-2)} \big(\frac{\Gamma(n)}{\Gamma(n/2)}\big)^{2/n}$ [Led00].

Simply taking the log:

$$\frac{n-2}{n} \log \Big(\int_{\mathbb{R}^n} |f|^q \,\mathrm{d}x \Big) \le \log \Big(C_n \int_{\mathbb{R}^n} \|\nabla f\|^2 \,\mathrm{d}x \Big).$$

Now, assuming that $\int f^2 \,\mathrm{d}x = 1$, applying **Jensen's inequality** with respect to the measure $f^2 \,\mathrm{d}x$ yields

$$\log \Big(\int_{\mathbb{R}^n} |f|^q \,\mathrm{d}x \Big) = \log \Big(\int_{\mathbb{R}^n} |f|^{q-2} f^2 \,\mathrm{d}x \Big) \ge \int_{\mathbb{R}^n} \log \big(|f|^{q-2} \big) f^2 \,\mathrm{d}x = \frac{q-2}{2} \int_{\mathbb{R}^n} \log(f^2) f^2 \,\mathrm{d}x,$$

hence we get the *logarithmic Sobolev inequality* for the Lebesgue measure

$$\int_{\mathbb{R}^n} f^2 \log f^2 \,\mathrm{d}x \le \frac{n}{2} \log \Big(C_n \int_{\mathbb{R}^n} \|\nabla f\|^2 \,\mathrm{d}x \Big). \tag{log-Sobolev$_{\mathrm{d}x}$}$$

The sharp constant yields $\frac{n}{2} \log(\frac{2}{n\pi e} \dots)$ on the right-hand side. By a change of variables $f^2 \,\mathrm{d}x \leftarrow f^2 \,\mathrm{d}\gamma_n$ with γ_n the standard Gaussian measure and **concavity of the log** (i.e. $\log x \le x - 1$),

$$\int_{\mathbb{R}^n} f^2 \log f^2 \,\mathrm{d}\gamma_n \le 2 \int_{\mathbb{R}^n} \|\nabla f\|^2 \,\mathrm{d}\gamma_n,$$

which gives the continuous embedding

$$\dot{W}^{1,2}_{\gamma_n}(\mathbb{R}^n) \hookrightarrow L^2_{\gamma_n} \log L^2_{\gamma_n}(\mathbb{R}^n). \tag{log-Sob. emb.$_2$}$$

Compare with (GNS emb.$_2$), which was

$$\dot{W}^{1,2}(\mathbb{R}^n) \hookrightarrow L^{2\frac{n}{n-2}}(\mathbb{R}^n) \xrightarrow[n\to\infty]{} L^2(\mathbb{R}^n). \tag{GNS emb.$_2$}$$

When the dimension goes to infinity, *a logarithmic gain remains*. Then by a change of variables $f^2 \leftarrow f$ and chain rule on ∇f,

$$\int_{\mathbb{R}^n} f \log f \,\mathrm{d}\gamma_n \le \frac{1}{2} \int_{\mathbb{R}^n} \frac{\|\nabla f\|^2}{f} \,\mathrm{d}\gamma_n = \frac{1}{2} \int_{\mathbb{R}^n} \|\nabla \log f\|^2 f \,\mathrm{d}\gamma_n.$$

The quantity on the left-hand side is sometimes written $\mathrm{Ent}_{\gamma_n}[f]$ in the literature, for instance in [BGL14], and the right-hand side is the Fisher information $I_{\gamma_n}[f]$ (or Dirichlet form), and this inequality can be written $\mathrm{Ent}_{\gamma_n}[f] \le \frac{1}{2} I_{\gamma_n}[f]$. This holds even if $\int f \,\mathrm{d}\gamma$ is not 1, with an additional term on the left-hand side. To express this condition with probability measures, just take $f = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_n}$ and get

$$\int_{\mathbb{R}^n} \rho \log \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_n} \le \frac{1}{2} \int_{\mathbb{R}^n} \rho \Big\| \nabla \log \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_n} \Big\|^2,$$

for all probability measures ρ, which is the *Stam–Gross log-Sobolev inequality* [Sta59; Gro75]

$$H(\rho \,|\, \gamma_n) \le \frac{1}{2} I(\rho \,|\, \gamma_n). \tag{log-Sobolev$_{\gamma_n}$}$$

Remark 2.15 (On the change of variable $f^2 \leftarrow f$) In the general case of a Dirichlet form $\mathcal{E}$, the *modified log-Sobolev inequality* [Van14, Theorem 3.20] reads

$$\mathrm{Ent}_\gamma[f] \le \frac{1}{2}\mathcal{E}(\log f, f) = \frac{1}{2}\int_{\mathbb{R}^n} \frac{\|\nabla f\|^2}{f}\,\mathrm{d}\gamma \qquad \text{(mod. log-Sobolev)}$$

compared to log-Sobolev:

$$\mathrm{Ent}_\gamma[f^2] \le 2\mathcal{E}(f,f) = 2\int_{\mathbb{R}^n} \|\nabla f\|^2\,\mathrm{d}\gamma$$

In our case where the Dirichlet form is $\int \|\nabla f\|^2\,\mathrm{d}\gamma$, that is, given in terms of a gradient that satisfies the chain rule, these two are equivalent [Van14]. △

Remark 2.16 (Log-Sobolev and Gaussian isoperimetric inequalities) [MV00] There is a notion of Gaussian isoperimetric inequality, Gaussian counterpart of the Lebesgue isoperimetric inequality we mentioned in Remark 2.13. △

More generally, we say that a measure γ satisfies a *logarithmic Sobolev inequality* with constant $K > 0$ if for all probability measures ρ

$$H(\rho\,|\,\gamma) \le \frac{1}{2K} I(\rho\,|\,\gamma). \qquad \text{(log-Sobolev)}$$

Straight away, we can link the Poincaré inequality to the logarithmic Sobolev inequality:

Remark 2.17 (Poincaré is linearized log-Sobolev)**.** For g smooth and such that $\int g\,\mathrm{d}\gamma = 0$,

$$H\big((1+\varepsilon g)\gamma\,|\,\gamma\big) = \frac{\varepsilon^2}{2}\|g\|^2_{L^2(\gamma)} + o(\varepsilon^2);$$
$$I\big((1+\varepsilon g)\gamma\,|\,\gamma\big) = \varepsilon^2\|\nabla g\|^2_{L^2(\gamma)} + o(\varepsilon^2). \qquad △$$

As a corollary, we get:

Theorem 2.18. (log-Sobolev)$_K$ *implies* (Poincaré)$_K$.

Remark 2.19 (Counterexample for the converse) The exponential measure $\frac{1}{2}e^{|x|}\,\mathrm{d}x$ on $\mathbb{R}$ satisfies a (Poincaré)$_1$ inequality, but does not satisfy any (log-Sobolev) inequality. More generally, the measure $\propto e^{|x|^\beta}\,\mathrm{d}x$ on $\mathbb{R}$ satisfies a (Poincaré) inequality if and only if $\beta \ge 1$, and a (log-Sobolev) inequality if and only if $\beta \ge 2$. △

Below is another answer to our Question 1.8 at the beginning:

Proposition 2.20 (Convergence in H under log-Sobolev)**.** *If γ satisfies* (log-Sobolev)$_K$*, then for ρ_t solution of* (Fokker–Planck)

$$H(\rho_t\,|\,\gamma) \le H(\rho_0\,|\,\gamma)e^{-t/2K}.$$

Actually, the converse also holds [Van14].

Proof of the direct implication. Grönwall, using that along the flow of (Fokker–Planck) by direct computation $\frac{\mathrm{d}}{\mathrm{d}t}H(\rho\,|\,e^{-V}) = -I(\rho\,|\,e^{-V})$. □

2.3.2. *When is log-Sobolev satisfied?* We just saw that having a (log-Sobolev) inequality satisfied is nice: it implies linear convergence to the equilibrium in terms of the relative entropy. It holds if γ is the standard Gaussian measure γ_n, see (log-Sobolev$_{\gamma_n}$). This is nice, but quite limited: this only concerns the very special case of the quadratic potential $V = \|\cdot\|^2/2$. More generally, to which class of measures can we extend this result?

Theorem 2.21 (Bakry–Émery condition [BÉ85])**.** *Consider a probability measure e^{-V} on some Riemannian manifold M. If*

$$\nabla^2 V + \mathrm{Ric}_M \ge K I_n, \qquad \text{(Bak–Ém)}$$

then e^{-V} satisfies (log-Sobolev)$_K$.

Lemma 2.22 (Holley–Stroock perturbation lemma)**.** *If V is of the form $V = V_0 + v$, where $v \in L^\infty$ and e^{-V_0} satisfies* (log-Sobolev)$_K$*, then e^{-V} satisfies* (log-Sobolev)$_{K\exp(\mathrm{osc}(v))}$*, where* $\mathrm{osc}(v) = \sup v - \inf v$.

By combining the Bakry–Émery condition with the Holley–Stroock lemma, one can generate a lot of probability measures satisfying a logarithmic Sobolev inequality [MV00].

If v is unbounded but satisfies $\int e^{\alpha\|\nabla v\|^2} e^{-V_0\,\mathrm{d}x}$ for α large enough, then V also satisfies a (log-Sobolev) inequality [Vil+09, Remark 21.5].

Remark 2.23 (Non-uniform convexity for V) Now, suppose that V behaves at infinity like $|x|^\alpha$ for $0 < \alpha < 2$. If $1 \leq \alpha < 2$, there is no (log-Sobolev) but there is a (Poincaré), hence the linear approach seems better suited at first; but it's okay, we can overcome the absence of (log-Sobolev), see [TV00], by compensating the degeneracy of the convexity using moments to localize the distribution function. We have

$$H(\rho\,|\,e^{-V}) \leq CI(\rho\,|\,e^{-V})^{1-\delta} M_s(\rho)^\delta,$$

where $M_s(\rho)$ is the moment of order $s > 2$ of ρ,

$$M_s(\rho) = \int_{\mathbb{R}^n} (1 + \|x\|^2)^{s/2}\,\mathrm{d}\rho(x) \qquad \text{and} \qquad \delta = \frac{2-\alpha}{2(2-\alpha)+s-2} \in (0, 1/2).$$

Combining this with a separate study of the time-behavior of moments, one can prove convergence to equilibrium with rate $O(t^{-\kappa})$ for any κ if the initial datum is rapidly decreasing. △

Remark 2.24 (Contraction techniques)**.** In his seminal work [Caf00], Caffarelli proved that the optimal transport map T between two log-concave measures γ_1 and γ_2 is Lipschitz, with a dimension-free bound L. The existence of such maps allows to transfer functional inequalities from one measure to the other. For instance, assuming γ_1 satisfies (log-Sobolev)$_K$ and $T_*\gamma_1 = \gamma_2$,

$$\int f^2 \log f^2 \gamma_2 = \int (f\circ T)^2 \log(f\circ T)^2 \gamma_1 \leq \frac{2}{K}\int \|\nabla(f\circ T)\|^2\gamma_1 \leq \frac{2L}{K}\int \|\nabla f\circ T\|^2\gamma_1 = \frac{2L}{K}\int f^2\log f^2\gamma_2,$$

hence γ_2 satisfies (log-Sobolev)$_{K/L}$. Since the standard Gaussian measure γ_n satisfies a (log-Sobolev) inequality, this recovers the (Bak–Ém) condition. Extensions of Caffarelli's contraction theorem have been proven since then: [CFJ15] generalizes it to perturbations of log-concave measures, [CFS24] to inverse powers of concave functions, and [DS24] to log-subharmonic measures.

Yet, in our case, it is not important that the map is optimal and any Lipschitz map will suffice. Kim and Milman [KM12] introduced a new construction of such transport maps via Langevin diffusion. This construction of a *flow map* does not coincide with the Brenier map in general [Tan21; LS22]. See [FMS24] for a detailed study of the Lipschitz properties of this flow map and [Kol11] for a review of generalizations and applications of Caffarelli's contraction result. See [DS24] for a generalization to log-subharmonic measures. See also [MS24; LS24]. △

Remark 2.25 (Tensorization)**.** If ρ_1 and ρ_2 satisfy (log-Sobolev) with constants K_1 and K_2, respectively, then $\rho_1 \otimes \rho_2$ satisfies (log-Sobolev) with constant $\min(K_1, K_2)$. △

Remark 2.26 (Intrinsic dimensional log-Sobolev [She24; ES24]) In this remark, we go back to the Lebesgue measure and consider the *entropy* H and the *Fisher information* I defined as

$$H(\rho) := H(\rho\,|\,\mathrm{d}x) = \int_{\mathbb{R}^n} \rho\log\rho \qquad \text{and} \qquad I(\rho) := I(\rho\,|\,\mathrm{d}x) = \int_{\mathbb{R}^n} \|\nabla \log\rho\|^2\rho,$$

respectively. With these notations, (log-Sobolev$_{\gamma_n}$) can be written [Sta59]

$$H(\rho) - H(\gamma_n) \leq \frac{1}{2}(I(\rho) - n), \tag{LS}$$

which can be seen for instance by using the change of reference measure formula [AGS14, Lemma 7.2]

$$H(\rho\,|\,\gamma_1) = H(\rho\,|\,\gamma_2) + \int_{\mathbb{R}^n} \log\left(\frac{\mathrm{d}\gamma_2}{\mathrm{d}\gamma_1}\right)\mathrm{d}\rho.$$

If instead, we stopped at (log-Sobolev$_{\mathrm{d}x}$), that is, before weakening the inequality using $\log x \leq x - 1$, we would get that the Lebesgue measure satisfies the (stronger) *dimensional log-Sobolev inequality* [Sta59; Car91; CC84]

$$H(\rho) - H(\gamma_n) \leq \frac{n}{2} \log\Big(\frac{I(\rho)}{n}\Big). \qquad \text{(dimensional LS)}$$

To get from (LS) to (dimensional LS), one can apply (LS) with the scaled measure $(T_\lambda)_* \rho$ where $\lambda > 0$ and $T_\lambda : x \mapsto \lambda x$. Optimizing over λ yields the (dimensional LS) inequality [Dem90]. See [ES24, Section 1.1.1.] for a formulation of (dimensional LS) over functions $f : \mathbb{R}^n \to \mathbb{R}$ and links to Beckner's inequality. When the Fisher information I is large, (dimensional LS) is exponentially better than (LS). Yet, it doesn't capture the intrinsic dimension of the measure ρ: suppose that $\rho = \tilde{\rho} \otimes \gamma_{n-k}$ where $\tilde{\rho} \in \mathcal{P}(\mathbb{R}^k)$ with $k \leq n$. Then, as n increases,

$$\frac{n}{2} \log\Big(\frac{I(\rho)}{n}\Big) = \frac{n}{2} \log\Big(\frac{I(\tilde{\rho}) + n - k}{n}\Big) \xrightarrow[n\to\infty]{} \frac{1}{2}(I(\tilde{\rho}) - n),$$

hence (dimensional LS) deteriorates to (LS) when the ambient dimension n increases, insensitive to the intrinsic dimension of ρ. Dembo [Dem90] showed that (dimensional LS) can be tightened as the (stronger) *intrinsic dimensional log-Sobolev inequality*

$$E(\rho) - E(\gamma_n) \leq \frac{1}{2} \log \det \mathcal{I}(\rho), \qquad \text{(intrinsic dimensional LS)}$$

where $\mathcal{I}$ is the *Fisher information matrix*

$$\mathcal{I}(\rho) = \int_{\mathbb{R}^n} (\nabla \log \rho)^{\otimes 2} \rho$$

Note that $I(\rho) = \operatorname{Tr} \mathcal{I}(\rho)$. From (dimensional LS), we tightened the bound by $\log \det M \leq n \log \frac{\operatorname{Tr} M}{n}$ which holds for any $n \times n$ positive definite matrix M. This inequality captures the intrinsic dimension of ρ, since each side of it behaves additively with respect to product measures: plugging in $\rho = \tilde{\rho} \otimes \gamma_{n-k}$ yields

$$H(\tilde{\rho}) - H(\gamma_k) \leq \frac{1}{2} \log \det \mathcal{I}(\tilde{\rho}),$$

as desired. To get from (dimensional LS) to (intrinsic dimensional LS), one can apply (dimensional LS) with the scaled measure $(T_\Lambda)_* \rho$ where Λ is a positive semidefinite matrix and $T_\Lambda : x \mapsto \Lambda x$. Optimizing over Λ yields the (intrinsic dimensional LS) inequality [ES24]. One could also imagine other families acting on measures that would tighten this inequality, or some other ones, as well. The optimization problem is sometimes too difficult to solve, but Eskenazis and Shenfeld [ES24] manage to do it for several inequalities. △

Remark 2.27 (Intrinsic dimensional convexity [She24]) First, let us recall some evolution equations on probability measures, starting with *optimal transport flows* (geodesics in the Otto geometry). These are flows (ρ_t, v_t) minimizing the classical Benamou–Brenier functional [BB00]

$$\frac{1}{2} \int_0^1 \int_\Omega \|v_t\|^2 \,\mathrm{d}\rho_t \,\mathrm{d}t,$$

while satisfying the continuity equation $\partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0$ between some ρ_0 and ρ_1. Such minimal flows satisfy the equation

$$\partial_t \rho_t + \operatorname{div}(\rho_t \nabla p_t) = 0, \qquad \text{where} \quad \partial_t p_t + \frac{1}{2} \|\nabla p_t\|^2 = 0. \qquad \text{(OT flow)}$$

Next, the *heat flow* is

$$\partial_t \rho_t - \frac{1}{2} \Delta \rho_t = 0. \qquad \text{(heat flow)}$$

Finally ,the *entropic interpolation flow* is the flow minimizing the functional

$$\int_0^1 \int_\Omega \Big(\frac{1}{2} \|v_t\|^2 + \frac{\varepsilon}{8} \|\nabla \log \rho_t\|^2\Big) \,\mathrm{d}\rho_t \,\mathrm{d}t$$

among flows satisfying the continuity equation. It yields

$$\partial_t \rho_t + \operatorname{div}(\rho_t \nabla p_t) = 0, \qquad \text{where} \quad \partial_t p_t + \frac{1}{2}\|\nabla p_t\|^2 + \frac{\varepsilon}{2}\frac{\Delta \rho_t^{1/2}}{\rho_t^{1/2}} = 0. \tag{entropic flow}$$

It is the dynamic formulation of entropic optimal transport and is also known as the *Schrödinger bridge problem* [CGP16; GLR17; CGP21]. It encapsulates both the (OT flow) and the (heat flow) in the limits $\varepsilon \to 0$ and $\varepsilon \to \infty$, respectively.

Now, back to our remark. McCann [McC97] showed that

$$\text{the map } H(\cdot \,|\, \mathrm{vol}) \text{ is convex along (OT flow).} \tag{DCvx}$$

This is also called *displacement convexity* (that is, geodesic convexity in the Otto geometry). Even more, on a Riemannian manifold M, this convexity is equivalent to M having nonnegative Ricci curvature, and this can be taken as the definition of nonnegative Ricci curvature on more general metric measure spaces (also written $\mathrm{CD}(K,\infty)$). But $H(\cdot \,|\, \mathrm{vol})$ is also convex along other flows: for instance, along the (entropic flow) [Léo17]. There is a stronger curvature condition which incorporates the effect of the dimension. Restricting to the flat case, this is the $\mathrm{CD}(0,n)$ condition of Bakry–Émery [BGL14]. Erbar, Kuwada, and Sturm [EKS15] showed that the $\mathrm{CD}(0,n)$ condition is equivalent to *dimensional displacement convexity*, that is,

$$\text{the map } e^{-H(\cdot \,|\, \mathrm{vol})/n} \text{ is concave along (OT flow).} \tag{dimensional DCvx}$$

It was then shown that the entropy is dimensional displacement convex along the (heat flow) [Cos85], then along the (entropic flow) [Rip19]. Shenfeld [She24] develop a new notion of *matrix* displacement convexity, which is stronger than dimensional displacement convexity (and thus than classical displacement convexity).

$$\text{the map ... is concave along (OT flow).} \tag{intrinsic dimensional DCvx}$$

△

3. Inequalities again and making sense of everything

3.1. Some more inequalities.

Let γ be the standard Gaussian measure. Then [Tal96]

$$W_2(\rho,\gamma) \le \sqrt{2H(\rho \,|\, \gamma)}.$$

More generally, we say that $\gamma = e^{-V}$ satisfies a *Talagrand inequality* with constant $K > 0$ if for all probability measures ρ

$$W_2(\rho,\gamma) \le \sqrt{\frac{2}{K} H(\rho \,|\, \gamma)}. \tag{Talagrand$_2$}$$

Theorem 3.1 ([OV00]). *One has the following chain of implications:*

(i) $(\text{log-Sobolev})_K$ *implies* $(\text{Talagrand}_2)_K$.

(ii) $(\text{Talagrand}_2)_K$ *implies* $(\text{Poincaré})_K$.

(iii) $\big[(\text{Talagrand}_2)_K$ *and* V *convex*$\big]$ *imply* $(\text{log-Sobolev})_{K/2}$.

In short:

$$(\text{log-Sobolev})_K \overset{(*)}{\Longrightarrow} (\text{Talagrand}_2)_K \Longrightarrow (\text{Poincaré})_K, \tag{3.1}$$

where $(*)$ *is an equivalence if* V *is convex (see [Vil+09, Theorem 22.21] for a more general statement).*

Remark 3.2 (Transport inequalities) The (Talagrand$_2$) inequality is a special case of the so-called *transport inequalities* [GL10], of the form

$$\alpha\big(\mathcal{T}_c(\rho,\gamma)\big) \le J(\rho \,|\, \gamma),$$

where $\mathcal{T}_c$ is the optimal transport cost with respect to a cost function c, $\alpha : [0,\infty) \to [0,\infty)$ is an increasing function such that $\alpha(0) = 0$, and J is some functional on probability measures. When J is the relative entropy H, one talks about *transport-entropy inequalities*. Among those:

(i) The (Talagrand$_2$) inequality corresponds to $\mathcal{T}_c = W_2^2$.
(ii) The (Talagrand$_1$) inequality corresponds to $\mathcal{T}_c = W_1$:

$$W_1(\rho,\gamma) \leq \sqrt{\frac{2}{K} H(\rho\,|\,\gamma)}. \tag{Talagrand$_1$}$$

Note that since by Jensen's inequality, $\mathcal{T}_d^2 \leq \mathcal{T}_{d^2}$, for a given metric d on the base space, (Talagrand$_1$) is always weaker than (Talagrand$_2$).
(iii) More generally, the (Talagrand$_p$) inequality corresponds to $\mathcal{T}_c = W_p^p$, and since $W_p \leq W_q$ for $p \leq q$, the Talagrand$_p$ inequalities become stronger as p increases.
(iv) The Csiszár–Kullback–(Pinsker) inequality corresponds to $\mathcal{T}_c = W_1$ where the metric on the base space is the *Hamming metric* $d_H(x,y) = \mathbf{1}_{x\neq y}$, since $\mathcal{T}_{d_H} = \|\cdot\|_{\mathrm{TV}}$.
(v) Define the *quadratic-linear cost* $c_{\mathrm{ql}}(x,y) = \min(d(x,y)^2, d(x,y))$. Then the *quadratic-linear transport inequality*

$$\mathcal{T}_{c_{\mathrm{ql}}}(\rho,\gamma) \leq C H(\rho\,|\,\gamma)$$

is equivalent to the (Poincaré) inequality [Vil+09, Theorem 22.25]

See also [Vil+09, Theorem 22.28] for the *generalized logarithmic Sobolev* and *generalized Poincaré* inequalities. When J is the relative Fisher information I, one talks about *transport-information inequalities* [Gui+09b; Gui+09a]. Among those,
(vi) The (W_2I) inequality corresponds to $\mathcal{T}_c = W_2^2$:

$$W_2(\rho,\gamma) \leq \sqrt{\frac{2}{K} I(\rho\,|\,\gamma)}. \tag{W_2I}$$

It is the analogue of (Talagrand$_2$).

$$\text{(log-Sobolev)} \overset{(*)}{\Longrightarrow} (W_2I) \Longrightarrow \text{(Poincaré)}$$

where $(*)$ is an equivalence if V is convex. See [Gui+09b] for the constants and more details, and also [Ped24] for instance. △

Remark 3.3 (Generalized Fisher information) It is possible to define a *generalized Fisher information* [Vil+09, Definition 20.6]

$$I_U(\rho\,|\,\gamma) = \int f U''(f) \|\nabla f\|^2 \,\mathrm{d}\gamma = \int_{\mathbb{R}^n} \frac{\|\nabla p(f)\|^2}{f}, \qquad f = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma},$$

where $U : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a convex C^2 function and $p(r) = rU'(r) - U(r)$. Corresponding variants of the (HWI) inequality can be found in [Vil+09, Theorem 20.10]. We recover the usual Fisher information by taking $U(r) = r \log r$. △

The implications in (3.1) are consequences of *HWI inequalities* [OV00], that mix the relative entropy H, the Wasserstein distance W_2 and the relative Fisher information I (hence the name HWI). A particular case of it was proven in [Ott01].

Theorem 3.4 (HWI inequality)**.** *Consider $\gamma = e^{-V}$ where $\nabla^2 V \geq KI$ (K not necessarily nonnegative). Then for any two probability distributions ρ_0 and ρ_1,*

$$H(\rho_0\,|\,\gamma) \leq H(\rho_1\,|\,\gamma) + W_2(\rho_0,\rho_1)\sqrt{I(\rho_0\,|\,\gamma)} - \frac{K}{2} W_2(\rho_0,\rho_1)^2. \tag{HWI}$$

In particular, if $K > 0$ then choosing $\rho_0 = \gamma$ yields

$$W_2(\rho,\gamma) \leq \sqrt{\frac{2}{K} H(\rho\,|\,\gamma)},$$

which is the Talagrand inequality (Talagrand$_2$) implied by (log-Sobolev), implied itself by the convexity of V by the Bakry–Émery condition (Bak–Ém), and choosing $\rho_1 = \gamma$ yields

$$H(\rho\,|\,\gamma) \leq W_2(\rho,\gamma)\sqrt{I(\rho\,|\,\gamma)} - \frac{K}{2} W_2(\rho,\gamma)^2$$

$$\leq \frac{1}{2\kappa} I(\rho \,|\, \gamma) + \frac{\kappa - K}{2} W_2(\rho, \gamma)^2 \qquad \text{for any } \kappa > 0 \text{ by Young's inequality}$$
$$= \frac{1}{2K} I(\rho \,|\, \gamma) \qquad \text{for } \kappa = K,$$

which is (log-Sobolev) implied by the convexity of V by the Bakry–Émery condition (Bak–Ém). Note that in any case, a (log-Sobolev) inequality is always satisfied up to a small second-order error term.

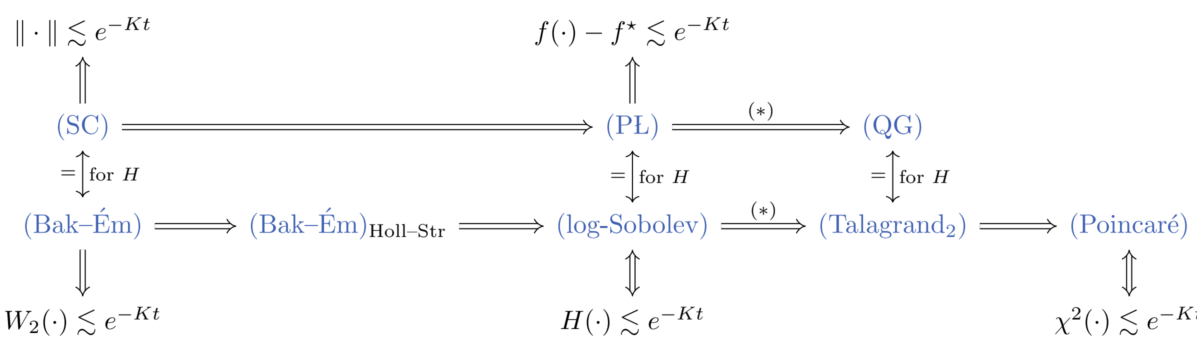
3.2. **Making sense of everything: interpretation via Otto calculus.** The last computation is exactly the same as in the Euclidean case (Section 1.1, Proposition 1.1), and for a good reason. In fact, conditions (Bak–Ém), (log-Sobolev) and (Talagrand$_2$) are just conditions (SC), (PŁ) and (QG) expressed in the Wasserstein–Otto geometry [Ott01] for the functional $H = H(\cdot \,|\, \gamma)$:

- (SC) Strong convexity of H in the Wasserstein geometry, that is, $\mathrm{Hess}_{W_2} H(\rho) \geq KI$ or displacement convexity, is the (Bak–Ém) condition;
- (PŁ) Since $\|\mathrm{grad}_{W_2} H(\rho)\|^2 = I(\rho \,|\, \gamma)$, the Polyak–Łojasiewicz condition for H is the (log-Sobolev) inequality, since the minimum $H^\star$ of H over the whole set of probability measures is zero;
- (QG) Quadratic growth for H in the Wasserstein geometry is directly (Talagrand$_2$).

See Ambrosio, Gigli, and Savaré [AGS14] for a rigorous in-depth study of this and many other things. To summarize the correspondences between the Euclidean and Wasserstein cases:

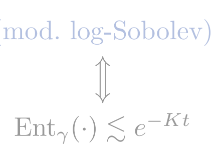
	In $\mathbb{R}^n$ for f		In $\mathcal{P}(\mathbb{R}^n)$ for $H(\cdot \,|\, e^{-V})$
(SC)	$\nabla^2 f \geq KI$	(Bak–Ém)	$\nabla^2 V \geq KI_n$
(PŁ)	$f(x) - f^\star \leq \frac{1}{2K}\|\nabla f(x)\|^2$	(log-Sobolev)	$H(\rho \,|\, e^{-V}) \leq \frac{1}{2K} I(\rho \,|\, e^{-V})$
(QG)	$\|x - x^\star\|^2 \leq \frac{2}{K}(f(x) - f^\star)$	(Talagrand$_2$)	$W_2(\rho, e^{-V}) \leq \sqrt{\frac{2}{K} H(\rho \,|\, e^{-V})}$
	?	(Poincaré)	$\chi^2(\rho \,|\, e^{-V}) \leq \frac{1}{K}\mathcal{E}_{e^{-V}}(\rho)$

And to sum up what we have so far,



where $(*)$ are equivalences if V (or f in Euclidean spaces) is convex. For the bottom equivalences, see [Van14]. Also, as in the Euclidean case, if (Talagrand$_2$) is satisfied, then linear convergence in H implies linear convergence in W_2. Since we always have (log-Sobolev) $\Longrightarrow$ (Talagrand$_2$), then (log-Sobolev) also implies linear convergence in terms of W_2.

Remark 3.5 (Modified log-Sobolev) In the more general case of Remark 2.15 (i.e. a more general Dirichlet form that doesn't necessarily satisfy the chain rule), (mod. log-Sobolev) and (log-Sobolev) are different. (mod. log-Sobolev) is actually equivalent to linear convergence in the entropy Ent_γ.



3.3. **Even more inequalities: other ϕ-divergences.** In Section 1.2 we used two different divergences to quantify how far our current measure ρ is from the target measure γ. Actually, one can define a family of relative entropy functionals, named ϕ-divergences, of the form

$$D_\phi : (\rho,\gamma) \longmapsto \int_{\mathbb{R}^n} \phi\Big(\frac{\rho}{\gamma}\Big)\gamma,$$

if $\rho \ll \gamma$, else ∞, with ϕ a convex function satisfying $\phi(1) = 0$. Those relative entropy functionals interpolate between the (classical) relative entropy ($\phi(h) = h\log h - h + 1$) and the $L^2(e^{-V})$ distance ($\phi(h) = (h-1)^2$). For each of these entropies one can prove log-Sobolev inequalities, which are stronger when the nonlinearity in the relative entropy is weaker (the strongest one is the $h\log h$ nonlinearity). Corresponding variants of the Holley–Stroock perturbation lemma and of the Csiszár–Kullback–(Pinsker) inequality are established in great generality in [Arn+98]. For instance [Tsy03]:

(i) for $\phi(h) = (h-1)^2$, this is the $L^2(e^{-V})$ distance;
(ii) for $\phi(h) = h^2 - 1$, this is the *chi-squared* χ^2 [Rig22; Che+20]

$$\chi^2(\rho\,|\,\gamma) = \int_{\mathbb{R}^n} \Big(\frac{\rho}{\gamma}\Big)^2\gamma - 1;$$

(iii) for $\phi(h) = |h-1|/2$, this is the total variation distance

$$\|\rho-\gamma\|_{\mathrm{TV}} = \sup_A |\rho(A) - \gamma(A)|,$$

which is also equal to $\frac{1}{2}\|\rho-\gamma\|_{L^1}$ for densities. It metrizes the strong convergence.
(iv) for $\phi(h) = (\sqrt{h}-1)^2$, this is the squared *Hellinger distance*

$$\mathrm{Hell}(\rho,\gamma)^2 = \int_{\mathbb{R}^n} \big(\sqrt{\rho}-\sqrt{\gamma}\big)^2.$$

It metrizes the strong convergence.
(v) for $\phi(h) = h\log h - h + 1$, this is the relative entropy H.

Proposition 3.6 (Inequalities between divergences [Tsy03])**.**

(i) H and χ^2:

$$H(\rho\,|\,\gamma) \le \log\big(\chi^2(\rho\,|\,\gamma) + 1\big) \le \chi^2(\rho\,|\,\gamma).$$

by definition, using Jensen for the first inequality and $\log x \le x-1$ for the second one. Actually, one has that

$$\begin{cases} H \approx \log\chi^2 & \text{when the distributions are far;} \\ H \approx \chi^2 & \text{when the distributions are close.} \end{cases}$$

(ii) Hellinger and TV:

$$\frac{1}{2}\,\mathrm{Hell}^2(\rho,\gamma) \le \|\rho-\gamma\|_{TV} \le \mathrm{Hell}(\rho,\gamma)\sqrt{1-\mathrm{Hell}^2(\rho,\gamma)/4} \qquad \text{(Le Cam)}$$

(iii) Hellinger and H:

$$\mathrm{Hell}^2(\rho,\gamma) \le H(\rho\,|\,\gamma) \qquad (3.2)$$

(iv) H and TV: Combining (Le Cam) and (3.2) would give $\|\rho-\gamma\|_{TV}^2 \le H(\rho\,|\,\gamma)$, but we can actually have a better constant:

$$\|\rho-\gamma\|_{TV}^2 \le \frac{1}{2}H(\rho\,|\,\gamma). \qquad \text{(Pinsker)}$$

See also [BV05] for a weighted version. When H is large (greater than 2), (Pinsker) is trivial, so one should rather use the Bretagnolle–Huber bound *[BH78; Can22; Tsy03]*

$$\|\rho-\gamma\|_{TV} \le \sqrt{1-e^{-H(\rho\,|\,\gamma)}} \le 1 - \frac{1}{2}e^{-H(\rho\,|\,\gamma)} \qquad \text{(BH)}$$

Summing up,

$$\|\rho-\gamma\|_{TV} \le \mathrm{Hell}(\rho,\gamma) \le \sqrt{H(\rho\,|\,\gamma)} \le \sqrt{\chi^2(\rho\,|\,\gamma)}.$$

References

[AGS14] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below". In: *Inventiones mathematicae* 195.2 (2014), pp. 289–391 (cit. on pp. 11, 15).

[Arn+98] Anton Arnold, Peter Markowich, Giuseppe Toscani, and Andreas Unterreiter. "On logarithmic Sobolev inequalities, Csiszar-Kullback inequalities, and the rate of convergence to equilibrium for Fokker-Planck type equations". In: *Preprint* (1998) (cit. on p. 16).

[BB00] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numerische Mathematik* 84.3 (2000), pp. 375–393 (cit. on p. 12).

[BÉ85] Dominique Bakry and Michel Émery. "Diffusions hypercontractives". In: *Séminaire de Probabilités XIX 1983/84: Proceedings*. Springer, 1985, pp. 177–206 (cit. on p. 10).

[BFP16] Annalisa Baldi, Bruno Franchi, and Pierre Pansu. "Gagliardo-Nirenberg inequalities for differential forms in Heisenberg groups". In: *Mathematische Annalen* 365 (2016), pp. 1633–1667 (cit. on p. 6).

[BFP22] Annalisa Baldi, Bruno Franchi, and Pierre Pansu. "Poincaré and Sobolev inequalities for differential forms in Heisenberg groups and contact manifolds". In: *Journal of the Institute of Mathematics of Jussieu* 21.3 (2022), pp. 869–920 (cit. on p. 6).

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 103. Springer, 2014 (cit. on pp. 5, 7, 9, 13).

[BH78] Jean Bretagnolle and Catherine Huber. "Estimation des densités: risque minimax". In: *Séminaire de probabilités de Strasbourg* 12 (1978), pp. 342–363 (cit. on p. 16).

[BL97] Sergey Bobkov and Michel Ledoux. "Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution". In: *Probability Theory and Related Fields* 107 (1997), pp. 383–400 (cit. on p. 6).

[BPV22] Raphaël Barboni, Gabriel Peyré, and François-Xavier Vialard. "On global convergence of ResNets: From finite to infinite width using linear parameterization". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 16385–16397 (cit. on p. 3).

[Bub+15] Sébastien Bubeck et al. "Convex optimization: Algorithms and complexity". In: *Foundations and Trends® in Machine Learning* 8.3-4 (2015), pp. 231–357 (cit. on p. 4).

[BV05] François Bolley and Cédric Villani. "Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities". In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 14. 3. 2005, pp. 331–352 (cit. on p. 16).

[Caf00] Luis A Caffarelli. "Monotonicity properties of optimal transportation and the FKG and related inequalities". In: *Communications in Mathematical Physics* 214 (2000), pp. 547–563 (cit. on p. 11).

[Can22] Clément L Canonne. "A short note on an inequality between KL and TV". In: (2022). arXiv: 2202.07198 (cit. on p. 16).

[Car91] Eric A Carlen. "Superadditivity of Fisher's information and logarithmic Sobolev inequalities". In: *Journal of Functional Analysis* 101.1 (1991), pp. 194–211 (cit. on p. 12).

[CC84] Max Costa and Thomas Cover. "On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (corresp.)" In: *IEEE Transactions on Information Theory* 30.6 (1984), pp. 837–839 (cit. on p. 12).

[CFJ15] Maria Colombo, Alessio Figalli, and Yash Jhaveri. "Lipschitz changes of variables between perturbations of log-concave measures". In: (2015). arXiv: 1510.03687 (cit. on p. 11).

[CFS24] Guillaume Carlier, Alessio Figalli, and Filippo Santambrogio. "On optimal transport maps between 1/d-concave densities". In: *arXiv preprint arXiv:2404.05456* (2024) (cit. on p. 11).

[CGP16] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint". In: *Journal of Optimization Theory and Applications* 169 (2016), pp. 671–691 (cit. on p. 13).

[CGP21] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. "Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrodinger bridge". In: *Siam Review* 63.2 (2021), pp. 249–313 (cit. on p. 13).

[Che+20] Sinho Chewi, Thibaut Le Gouic, Chen Lu, Tyler Maunu, Philippe Rigollet, and Austin Stromme. "Exponential ergodicity of mirror-Langevin diffusions". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 19573–19585 (cit. on pp. 7, 16).

[Cos85] Max Costa. "A new entropy power inequality". In: *IEEE Transactions on Information Theory* 31.6 (1985), pp. 751–760 (cit. on p. 13).

[Csi63] Imre Csiszár. "Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten". In: *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* 8.1-2 (1963), pp. 85–108 (cit. on p. 5).

[Dem90] Amir Dembo. "Information inequalities and uncertainty principles". In: *Dept. Statistics, Stanford Univ., Stanford, CA, Tech. Rep* 75 (1990) (cit. on p. 12).

[DMP24] Lorenzo Dello Schiavo, Jan Maas, and Francesco Pedrotti. "Local conditions for global convergence of gradient flows and proximal point sequences in metric spaces". In: *Transactions of the American Mathematical Society* 377.06 (2024), pp. 3779–3804 (cit. on pp. 2, 3).

[Dot] Paige Dote. *Sobolev and Isoperimetric Inequalities on Riemannian Manifolds.* URL: https://web.mit.edu/paigeb/www/994paper.pdf (cit. on p. 8).

[DS24] Guido De Philippis and Yair Shenfeld. *Optimal transport maps, majorization, and log-subharmonic measures.* 2024. arXiv: 2411.12109 (cit. on p. 11).

[EKS15] Matthias Erbar, Kazumasa Kuwada, and Karl-Theodor Sturm. "On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces". In: *Inventiones mathematicae* 201.3 (2015), pp. 993–1071 (cit. on p. 13).

[ES24] Alexandros Eskenazis and Yair Shenfeld. "Intrinsic dimensional functional inequalities on model spaces". In: *Journal of Functional Analysis* 286.7 (2024), p. 110338 (cit. on pp. 11, 12).

[FG21] Spencer Frei and Quanquan Gu. "Proxy convexity: A unified framework for the analysis of neural networks trained by gradient descent". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 7937–7949 (cit. on p. 3).

[FMS24] Max Fathi, Dan Mikulincer, and Yair Shenfeld. "Transportation onto log-Lipschitz perturbations". In: *Calculus of Variations and Partial Differential Equations* 63.3 (2024), p. 61 (cit. on p. 11).

[Gar15] Guillaume Garrigos. "Descent dynamical systems and algorithms for tame optimization and multi-objective problems." PhD thesis. Université de Montpellier; Universidad Tecnica Federico Santa Maria, 2015 (cit. on p. 3).

[GG23] Guillaume Garrigos and Robert M Gower. "Handbook of convergence theorems for (stochastic) gradient methods". In: *arXiv preprint arXiv:2301.11235* (2023) (cit. on pp. 2, 3).

[GL10] Nathael Gozlan and Christian Léonard. "Transport inequalities. A survey". In: (2010). arXiv: 1003.3852 (cit. on p. 13).

[GLR17] Ivan Gentil, Christian Léonard, and Luigia Ripani. "About the analogy between optimal transport and minimal entropy". In: *Annales de la Faculté des sciences de Toulouse: Mathématiques.* Vol. 26. 3. 2017, pp. 569–600 (cit. on p. 13).

[GM73] S Gallot and D Meyer. "Sur la premiere valeur propre du p-spectre pour les variétésa opérateur de courbure positif". In: *CR Acad. Sci. Paris Sér. AB* 276 (1973), A1619–A1621 (cit. on p. 6).

[Gro75] Leonard Gross. "Logarithmic sobolev inequalities". In: *American Journal of Mathematics* 97.4 (1975), pp. 1061–1083 (cit. on pp. 8, 9).

[Gui+09a] Arnaud Guillin, Christian Léonard, Feng-Yu Wang, and Liming Wu. "Transportation-information inequalities for Markov processes (II): relations with other functional inequalities". In: (2009). arXiv: 0902.2101 (cit. on p. 14).

[Gui+09b] Arnaud Guillin, Christian Léonard, Liming Wu, and Nian Yao. "Transportation-information inequalities for Markov processes". In: *Probability theory and related fields* 144 (2009), pp. 669–695 (cit. on p. 14).

[HM24] Shouhei Honda and Andrea Mondino. "Poincar\'e inequality for one forms on four manifolds with bounded Ricci curvature". In: *arXiv preprint arXiv:2405.19168* (2024) (cit. on p. 6).

[KM12] Young-Heon Kim and Emanuel Milman. "A generalization of Caffarelli's contraction theorem via (reverse) heat flow". In: *Mathematische Annalen* 354.3 (2012), pp. 827–862 (cit. on p. 11).

[KNS16] Hamed Karimi, Julie Nutini, and Mark Schmidt. "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition". In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16*. Springer. 2016, pp. 795–811 (cit. on pp. 2, 3).

[Kol11] Alexander V Kolesnikov. "Mass transportation and contractions". In: (2011). arXiv: 1103.1479 (cit. on p. 11).

[Kul59] Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1959 (cit. on p. 5).

[Kur98] Krzysztof Kurdyka. "On gradients of functions definable in o-minimal structures". In: *Annales de l'institut Fourier*. Vol. 48. 3. 1998, pp. 769–783 (cit. on p. 3).

[LDZ24] Feng-Yi Liao, Lijun Ding, and Yang Zheng. "Error bounds, PL condition, and quadratic growth for weakly convex functions, and linear convergences of proximal point methods". In: *6th Annual Learning for Dynamics & Control Conference*. PMLR. 2024, pp. 993–1005 (cit. on pp. 2, 3).

[Led00] Michel Ledoux. *Logarithmic Sobolev Inequalities*. 2000. URL: https://www.math.univ-toulouse.fr/~ledoux/Logsobwpause.pdf (cit. on pp. 5, 9).

[Léo17] Christian Léonard. "On the convexity of the entropy along entropic interpolations". In: *preprint* (2017) (cit. on p. 13).

[Łoj63] Stanislaw Łojasiewicz. "A topological property of real analytic subsets". In: *Coll. du CNRS, Les équations aux dérivées partielles* 117.87-89 (1963), p. 2 (cit. on pp. 2, 3, 15).

[LS22] Hugo Lavenant and Filippo Santambrogio. "The flow map of the Fokker–Planck equation does not provide optimal transport". In: *Applied Mathematics Letters* 133 (2022), p. 108225 (cit. on p. 11).

[LS24] Pablo López-Rivera and Yair Shenfeld. "The Poisson transport map". In: (2024). arXiv: 2407.02359 (cit. on p. 11).

[LZB22] Chaoyue Liu, Libin Zhu, and Mikhail Belkin. "Loss landscapes and optimization in over-parameterized non-linear systems and neural networks". In: *Applied and Computational Harmonic Analysis* 59 (2022), pp. 85–116 (cit. on p. 3).

[McC97] Robert J McCann. "A convexity principle for interacting gases". In: *Advances in mathematics* 128.1 (1997), pp. 153–179 (cit. on p. 13).

[MS24] Dan Mikulincer and Yair Shenfeld. "The Brownian transport map". In: *Probability Theory and Related Fields* (2024), pp. 1–66 (cit. on p. 11).

[MV00] Peter A Markowich and Cédric Villani. "On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis". In: *Mat. Contemp* 19 (2000), pp. 1–29 (cit. on pp. 4, 5, 10, 11).

[Ott01] Felix Otto. "The geometry of dissipative evolution equations: the porous medium equation". In: (2001) (cit. on pp. 14, 15).

[OV00] Felix Otto and Cédric Villani. "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality". In: *Journal of Functional Analysis* 173.2 (2000), pp. 361–400 (cit. on pp. 13, 14).

[Ped24] Francesco Pedrotti. "Functional inequalities and convergence of stochastic processes". PhD thesis. 2024 (cit. on p. 14).

[Pin64] Mark S Pinsker. "Information and information stability of random variables and processes". In: *Holden-Day* (1964) (cit. on p. 5).

[Pol63] Boris T Polyak. "Gradient methods for solving equations and inequalities". In: *USSR Computational Mathematics and Mathematical Physics* 4.6 (1963), pp. 17–32 (cit. on pp. 2, 15).

[Rig22] Philippe Rigollet. *An optimization perspective on sampling using optimal transport*. 2022. URL: https://theodumont.github.io/assets/pdf/rigollet_ot.pdf (cit. on pp. 5, 16).

[Rip19] Luigia Ripani. "Convexity and regularity properties for entropic interpolations". In: *Journal of Functional Analysis* 277.2 (2019), pp. 368–391 (cit. on p. 13).

[She24] Yair Shenfeld. "Matrix displacement convexity along density flows". In: *Archive for Rational Mechanics and Analysis* 248.5 (2024), p. 74 (cit. on pp. 11–13).

[Sob38] S Soboleff. "Sur un théorème d'analyse fonctionnelle". In: *Matematicheskii Sbornik* 46.3 (1938), pp. 471–497 (cit. on p. 7).

[Sta59] Aart J Stam. "Some inequalities satisfied by the quantities of information of Fisher and Shannon". In: *Information and Control* 2.2 (1959), pp. 101–112 (cit. on pp. 9, 11, 12).

[Str24] Austin Stromme. "Asymptotic log-Sobolev constants and the Polyak-Łojasiewicz gradient domination condition". Séminaire Parisien d'Optimisation. 2024 (cit. on pp. 2, 3).

[Tal96] Michel Talagrand. "Transportation cost for Gaussian and other product measures". In: *Geometric & Functional Analysis GAFA* 6.3 (1996), pp. 587–600 (cit. on p. 13).

[Tan21] Anastasiya Tanana. "Comparison of transport map generated by heat flow interpolation and the optimal transport Brenier map". In: *Communications in Contemporary Mathematics* 23.06 (2021), p. 2050025 (cit. on p. 11).

[Tsy03] Alexandre B Tsybakov. *Introduction à l'estimation non paramétrique*. Vol. 41. Springer Science & Business Media, 2003 (cit. on p. 16).

[TV00] Guiseppe Toscani and Cédric Villani. "On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds". In: *Journal of Statistical Physics* 98 (2000), pp. 1279–1309 (cit. on p. 11).

[Van14] Ramon Van Handel. "Probability in high dimension". In: *Lecture Notes (Princeton University)* 2.3 (2014), pp. 2–3 (cit. on pp. 6, 10, 15).

[Vil+09] Cédric Villani et al. *Optimal transport: old and new*. Vol. 338. Springer, 2009 (cit. on pp. 5, 6, 11, 13, 14).

[Vil03] Cédric Villani. *Topics in optimal transportation*. Vol. 58. American Mathematical Soc., 2003 (cit. on p. 5).