Isoperimetric inequalities in high-dimensional convex sets

Lecture notes by Bo'az Klartag and Joseph Lehec*

Institut Henri Poincaré (IHP), Paris, May 21–24, 2024

Abstract

These are lecture notes focusing on recent progress towards Bourgain's slicing problem and the isoperimetric conjecture proposed by Kannan, Lovasz and Simonovits (KLS).

Contents

Acknowledgement. We would like to thank the Institut Henri Poincaré for hosting the series of lectures. These notes are a direct outcome of those lectures. We are grateful to Esther Bou Dagher, Andreas Malliaris, and especially Cyril Roberto for the impeccable organization of this event. We also thank the anonymous referees for their careful reading of the manuscript and accurate comments.

*Sections 1,3,4,5,9 correspond to lectures by B.K. while Sections 2,6,7,8 correspond to lectures by J.L.

1 The Poincaré inequality

Even if we were not hosted by an institution that honors Poincaré, a good starting point for these lectures would be the mathematical inequality that carries his name. It was published by Poincaré in 1892–1896 in the case where the dimension is 2 or 3, and the measure μ is the uniform probability measure on a convex body K.

Recall that an absolutely continuous measure μ in $\mathbb{R}^n$ is *log-concave* if its density ρ is log-concave, namely

$$\rho(\lambda x + (1-\lambda)y) \geq \rho(x)^{\lambda}\rho(y)^{1-\lambda} \qquad (x, y \in \mathbb{R}^n, 0 < \lambda < 1). \tag{1}$$

In general, a Borel measure μ in $\mathbb{R}^n$ is log-concave if it is supported in an affine subspace and has a log-concave density in this subspace. The uniform probability measure on a convex body is log-concave, as well as all Gaussian measures.

Theorem 1 ("The Poincaré inequality")**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body, let μ be a log-concave probability measure on K. Then for any C^1-smooth function $f : K \to \mathbb{R}$ with $\int_K f d\mu = 0$,*

$$\int_K f^2 d\mu \leq C_P(\mu) \cdot \int_K |\nabla f|^2 \, d\mu \tag{2}$$

where $C_P(\mu) \leq C_n \cdot Diam^2(K)$, and $C_n > 0$ depends only on the dimension n.

Here $Diam(K) = \sup_{x,y\in K} |x-y|$ is the diameter of K and $|\cdot|$ is the standard Euclidean norm in $\mathbb{R}^n$. Intuitively, the inequality says that if f does not vary too wildly locally, i.e. controlled gradient, then it does not vary too much globally, i.e. bounded variance.

For a historical account of the Poincaré inequality, see Allaire [2]. The Poincaré constant $C_P(\mu)$ of the probability measure μ is defined as the smallest number for which (2) is valid for all C^1-smooth functions f with $\int f d\mu = 0$.

The quantity $1/C_P(\mu)$ is often referred to as the *spectral gap* of μ, for reasons to be explained. In 1960, Payne and Weinberger [73] found that for any n, the best possible value of the supposedly-dimensional constant C_n is in fact

$$C_n = \frac{1}{\pi^2},$$

which does not depend on the dimension. We proceed with an adaptation of the original proof by Poincaré, a proof which does not yield the optimal (in)dependence on the dimension, yet it suffices for some purposes.

Proof of Theorem 1. Passing to a subspace if necessary, we may assume that the probability measure μ is absolutely continuous with a log-concave density $\rho : \mathbb{R}^n \to [0, \infty)$, which vanishes outside K. We express the variance as a double integral and use the fundamental theorem of calculus:

$$\begin{aligned}\int_K f^2 d\mu &= \frac{1}{2}\int_K\int_K |f(y)-f(x)|^2\,\mu(dx)\mu(dy)\\ &= \frac{1}{2}\int_K\int_K \left|\int_0^1 \nabla f((1-t)x+ty)\cdot(y-x)dt\right|^2\,\mu(dx)\mu(dy)\\ &\le \frac{Diam^2(K)}{2}\int_K\int_K\int_0^1 |\nabla f((1-t)x+ty)|^2\,\rho(x)\rho(y)\,dtdxdy,\end{aligned}$$

where we used the inequality $|y-x|\le Diam(K)$. Let us show that for any $0\le t\le 1$,

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |\nabla f((1-t)x+ty)|^2\,\rho(x)\rho(y)\,dxdy \le C_{n,t}\int_{\mathbb{R}^n}|\nabla f|^2\,d\mu, \tag{3}$$

We integrate over $\mathbb{R}^n$ now, but recall that the density ρ vanishes outside K, so this does not make a difference. Our goal is to replace the product $\rho(x)\rho(y)$ in (3) by some expression involving $\rho((1-t)x+ty)$ and then apply a linear change of variables. Log-concavity will be handy here. We split the argument into two cases. If $t\approx 1/2$, then we will use the inequality

$$\min\{\rho(x),\rho(y)\}\le\rho((1-t)x+ty)$$

that follows from the definition (1) of log-concavity. It implies that

$$\rho(x)\rho(y)\le\rho((1-t)x+ty)\cdot\max\{\rho(x),\rho(y)\}\le\rho((1-t)x+ty)\cdot[\rho(x)+\rho(y)].$$

Thus the integral in (3) is at most

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|\nabla f((1-t)x+ty)|^2\,\rho((1-t)x+ty)\cdot[\rho(x)+\rho(y)]dxdy \qquad \text{“}u=(1-t)x+ty\text{''}$$

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|\nabla f(u)|^2\rho(u)\rho(x)\frac{du}{t^n}dx+\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|\nabla f(u)|^2\rho(u)\rho(y)\frac{du}{(1-t)^n}dy$$

$$=\left[\frac{1}{t^n}+\frac{1}{(1-t)^n}\right]\int_{\mathbb{R}^n}|\nabla f|^2 d\mu.$$

In the case where t is not too close to $1/2$ we will use the inequality

$$\rho(x)\rho(y)\le\rho((1-t)x+ty)\rho(tx+(1-t)y)$$

and change variables linearly via

$$u=(1-t)x+ty,\qquad v=tx+(1-t)y.$$

Since $du_j\wedge dv_j=[(1-t)^2-t^2]dx_j\wedge dy_j$ for $j=1,\dots,n$, the integral in (3) is bounded by

$$\begin{aligned}&\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|\nabla f((1-t)x+ty)|^2\,\rho((1-t)x+ty)\rho(tx+(1-t)y)dxdy\\ &=\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|\nabla f(u)|^2\rho(u)\rho(v)\frac{dudv}{|t^2-(1-t)^2|^n}=\frac{1}{|1-2t|^n}\int_{\mathbb{R}^n}|\nabla f|^2 d\mu.\end{aligned}$$

Thus the Poincaré inequality follows with

$$C_n \leq \frac{1}{2} \int_0^1 \min \left\{ \frac{1}{t^n} + \frac{1}{(1-t)^n}, \frac{1}{|1-2t|^n} \right\} dt \leq C \cdot \frac{3^n}{n},$$

for some universal constant $C > 0$, where we separately consider the contribution of the intervals $[0, 1/3], [1/3, 2/3], [2/3, 1]$ to the integral. □

Throughout these lectures, we write $C, c, \widetilde{C}, \widetilde{c}, \bar{C}$ etc. to denote various positive universal constants whose value may change from one line to the next. Consider the case where μ is the uniform probability measure on a domain $K \subseteq \mathbb{R}^n$. Its Poincaré constant, sometimes denoted also by $C_P(K)$, measures the *conductance* of K. It is large when K has a bottleneck.

Intuitively, it seems that convexity assumptions rule out many types of bottlenecks, possibly in high dimensions as well. Can we describe the Poincaré constant in terms of simple geometric characteristics of $K \subseteq \mathbb{R}^n$, under convexity assumptions?

Conjecture 2 (Kannan-Lovász-Simonovits [45]). *For any log-concave probability measure μ on $\mathbb{R}^n$,*

$$\|\mathrm{Cov}(\mu)\|_{op} \leq C_P(\mu) \leq C \cdot \|\mathrm{Cov}(\mu)\|_{op} \tag{4}$$

where $C > 0$ is a universal constant.

Here $\|A\|_{op}$ is the operator norm of the symmetric matrix $A \in \mathbb{R}^{n \times n}$, i.e., its maximal eigenvalue in absolute value, and $\mathrm{Cov}(\mu) \in \mathbb{R}^{n \times n}$ is the inertia matrix or the *covariance matrix* of μ. The i, j entry of the matrix $\mathrm{Cov}(\mu)$ is

$$\int_{\mathbb{R}^n} x_i x_j \, \mu(dx) - \int_{\mathbb{R}^n} x_i \, \mu(dx) \int_{\mathbb{R}^n} x_j \, \mu(dx).$$

The covariance matrix is a symmetric, positive semi-definite matrix. If X is a random vector with law μ and density ρ, we write $C_P(X) = C_P(\mu) = C_P(\rho)$ and $\mathrm{Cov}(X) = \mathrm{Cov}(\mu) = \mathrm{Cov}(\rho)$. With this notation, the Poincaré inequality states that for any C^1-smooth function f,

$$\mathrm{Var}(f(X)) \leq C_P(X) \cdot \mathbb{E}|\nabla f(X)|^2.$$

Originally the conjecture by Kannan, Lovász and Simonovits [45] was formulated in terms of a Cheeger inequality rather than a Poincaré inequality, but the two formulations turn out to be equivalent. We shall return to this in the next section. For various perspectives on the KLS conjecture, we refer the reader to the monographs by Artstein-Avidan, Giannopoulos and Milman [4] and by Brazitikos, Giannopoulos, Valettas and Vritsiou [25], as well as to the survey papers by Ball [8] and by Lee and Vempala [62].

We note that the left-hand side inequality in (4) is a trivial fact: for any linear functional $f_\theta(x) = x \cdot \theta$ with $\theta \in S^{n-1} = \{x \in \mathbb{R}^n \, ; \, |x| = 1\}$,

$$\mathrm{Cov}(X)\theta \cdot \theta = \mathrm{Var}(f_\theta(X)) \leq C_P(X) \cdot \mathbb{E}|\nabla f_\theta(X)|^2 = C_P(X),$$

and (4) follows by taking the supremum over all $\theta \in S^{n-1}$. Thus the KLS conjecture suggests that in the log-concave case, the Poincaré inequality is saturated by linear functions, up to a universal constant.

Exercise 1 (Tensorization)**.** For μ, ν probability measures on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively,

$$C_P(\mu \otimes \nu) = \max(C_P(\mu), C_P(\nu)).$$

Here are examples of log-concave measures for which we can compute the Poincaré constant.

1. Consider the one-dimensional case, where X is a random variable that is distributed uniformly in some interval of length L. Then,

$$\mathrm{Var}(X) = \frac{L^2}{12} \qquad \text{and} \qquad C_P(X) = \frac{L^2}{\pi^2},$$

   with the extremal function for the Poincaré inequality on $[0, \pi]$ being $f(x) = \cos x$.

2. Consider the case where X is distributed uniformly in $K = [0,1]^n$. In this case,

$$Diam(K) = \sqrt{n}$$

   while by the tensorization property of the Poincaré constant (see the exercise above)

$$C_P(X) = \frac{1}{\pi^2}$$

   and

$$\mathrm{Cov}(X) = \frac{1}{12} \cdot \mathrm{Id}.$$

   We thus see that the diameter bound for the Poincaré constant is rather weak in high dimensions, even with the optimal, dimension-independent constant.

3. Suppose that X is distributed uniformly in a Euclidean ball. The Euclidean unit ball $B^n = \{x \in \mathbb{R}^n \, ; \, |x| \leq 1\}$ has volume

$$\frac{\pi^{n/2}}{\Gamma(1 + n/2)} = \left(\frac{\sqrt{2\pi e} + o(1)}{\sqrt{n}} \right)^n,$$

   which is a rather small number in high dimensions. In order to normalize the volume (or the covariance, or the Poincaré constant), we had better look at the random vector X that is distributed uniformly in a Euclidean ball $K = \sqrt{n} \cdot B^n$. In this case,

$$Diam(K) = 2\sqrt{n}, \qquad \mathrm{Cov}(X) = \frac{n}{n+2} \cdot \mathrm{Id}.$$

   The Poincaré constant of X may be described using Bessel functions, and it has the order of magnitude of a universal constant, in accordance with the KLS conjecture. The Szegö-Weinberger inequality [80, 83] states that among all uniform distributions on domains in $\mathbb{R}^n$ of fixed volume, the Poincaré constant is minimized for a Euclidean ball.

4. Next we discuss the case where X is a standard Gaussian random vector in $\mathbb{R}^n$. Here,

$$\mathrm{Cov}(X) = \mathrm{Id} \qquad \text{and} \qquad C_P(X) = 1.$$

Thus the Poincaré inequality in the Gaussian case is precisely saturated by linear functions. Furthermore, by considering Hermite polynomials one can show the following: In the Gaussian case, a function nearly saturates the Poincaré inequality if and only if it is nearly a low-degree polynomial. Indeed, in one direction, if f is a polynomial of degree at most d in n real variables then we can reverse the Poincaré inequality as follows:

$$\mathbb{E}|\nabla f(X)|^2 \le d \cdot \mathrm{Var}(f(X)).$$

In the other direction, if f is a smooth function with

$$\mathbb{E}|\nabla f(X)|^2 \le R \cdot \mathrm{Var}(f(X))$$

then the function f may be approximated by a polynomial of bounded degree: For any $d \ge 0$ there exists a polynomial P of degree at most d such that

$$\mathbb{E}|(f - P)(X)|^2 \le \frac{R}{d+1} \cdot \mathrm{Var}(f(X)).$$

In fact, this polynomial P is obtained by truncating the Hermite expansion of f.

5. Let us work in $\mathbb{C}^n$ and consider the probability measure μ on $\mathbb{C}^n$ with density

$$\prod_{j=1}^{n} \frac{e^{-|z_j|}}{2\pi}.$$

The measure μ is a log-concave probability measure on $\mathbb{C}^n$. Its covariance matrix is

$$\mathrm{Cov}(\mu) = 3 \cdot \mathrm{Id}$$

and its Poincaré constant has the order of magnitude of a universal constant, in accordance with the KLS conjecture.

The density of μ decays expoentially at infinity. Exponentially, but not faster; any log-concave probability density decays exponentially at infinity, yet the Gaussian density decays even faster. This reflects on spectral properties. In the exponential case there are functions that nearly saturate the Poincaré inequality, and they do not necessarily resemble low-degree polynomials. For instance:

Claim: For any holomorphic function $f : \mathbb{C}^n \to \mathbb{C}$ with $f \in L^2(\mu)$ and $\int f d\mu = 0$ (or equivalently, with $f(0) = 0$), the Rayleigh quotient satisfies

$$\frac{1}{3} \le \frac{\int_{\mathbb{C}^n} |\nabla f|^2 \, d\mu}{\int_{\mathbb{C}^n} |f|^2 \, d\mu} \le \frac{1}{2}. \tag{5}$$

Here is a proof for $n = 1$, which can be easily generalized for any dimension. It suffices to check the validity of (5) for monomials z^k, because of orthogonality relations. If $f(z) = z^k$ with $k \geq 1$ then,

$$\|f\|_{L^2(\mu)}^2 = (2k+1)!$$

while

$$\|f'\|_{L^2(\mu)}^2 = k^2(2k-1)!$$

The ratio between the two is always between 4 and 6. We remark that by considering the real part of f, we see that (5) holds true for any pluri-harmonic function f, and in particular, when $n = 1$ the relation (5) holds true for any harmonic function $f : \mathbb{R}^2 \to \mathbb{R}$ (thanks to A. Eskenazis for suggesting to add this remark).

Exercise 2 (Subbaditivity)**.** For two independent random vectors X and Y in $\mathbb{R}^n$,

$$C_P(X+Y) \leq C_P(X) + C_P(Y).$$

1.1 Applications

Poincaré's original motivation for his inequality was related to analysis of partial differential equations such as the *heat equation*. The motivation of Kannan, Lovász and Simonovits in the 1990s came from algorithms based on Markov chains (MCMC) for sampling and for estimating the volume of a high-dimensional convex body. Such tasks appear in linear programming. Another motivation for this research direction, that was put forth by Ball in the early 2000s and later jointly with Nguyen [9], was the relation to Bourgain's slicing problem discussed below. There are models in probability and statistical physics for which log-concavity and Poincaré inequalities are relevant. Let us describe here another application, related to the *Central Limit Theorem for Convex Sets* [48] from 2006.

A random vector X in $\mathbb{R}^n$ is *isotropic* or *normalized* if $\mathbb{E}X = 0$ and

$$\mathrm{Cov}(X) = \mathrm{Id}.$$

Any random vector with finite second moments can be made isotropic by applying an affine-linear transformation. The relation between Gaussian approximation and the Poincaré constant stems from the following:

(i) The Poincaré inequality with $f(x) = |x|$ yields $\mathrm{Var}(|X|) \leq C_P(X)$. Thus most of the mass of an isotropic random vector X is contained in spherical shell

$$\left\{ x \in \mathbb{R}^n \, ; \, \sqrt{n} - 3\sqrt{C_P(X)} \leq |x| \leq \sqrt{n} + 3\sqrt{C_P(X)} \right\},$$

whose width has the order of magnitude of the square root of the Poincaré constant.

(ii) Gaussian approximation principle (Sudakov [78], Diaconis-Freedman [34]): When most of the mass of the isotropic random vector X is contained in a thin spherical shell, we have *approximately Gaussian marginals*.

The following theorem is the current state of the art on Gaussian approximation under Poincaré inequality. We write σ_{n-1} for the uniform probability measure on the unit sphere S^{n-1}.

Theorem 3 (Bobkov, Chistyakov, Götze [17, Proposition 17.5.1])**.** *Let X be an isotropic random vector in $\mathbb{R}^n$. Then there exists a subset $\Theta \subseteq S^{n-1}$ with $\sigma_{n-1}(\Theta) \geq 9/10$ such that any $\theta \in \Theta$,*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(X \cdot \theta \leq t) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2} ds \right| \leq \frac{C \log n}{n} \cdot C_P(X),$$

where $C > 0$ is a universal constant.

We do not know whether the logarithmic factor in Theorem 3 is necessary. It is currently known that $C_P(X) \leq C \cdot \log n$ for an isotropic, log-concave random vector X in $\mathbb{R}^n$, see [54]. Consequently Theorem 3 yields good error estimates in the Central Limit Theorem for Convex sets, and more generally for log-concave measures.

If all we know about the Poincaré constant is the diameter bound, then even in the case of the cube we would be off by a factor of n, and we would not obtain any non-trivial bound for the Central Limit Theorem for Convex sets. Thus in high dimensions it is necessary to refine the diameter bound, as suggested in the KLS conjecture.

What techniques can we use to this end, techniques that go beyond change of variables, Fubini theorem, and the Cauchy-Schwartz inequality used above? High-dimensional convex geometry is a playground for various geometric and analytic ideas that transcend the field of convexity. Any list of approaches that have proven useful to convexity must include convex localization, optimal transport, curvature and the Bochner formula, semigroup tools, geometric measure theory, stochastic localization and complex analysis. In these lectures we explore only some of these directions.

1.2 1D log-concave distributions

Before going on to study methods for high dimensions, let us briefly discuss the one-dimensional case. What do log-concave densities look like in one dimension?

Proposition 4 ("How to think on 1D log-concave random variables")**.** *Let $X \in \mathbb{R}$ be a log-concave random variable with density ρ which is isotropic. Then for any $x \in \mathbb{R}$,*

$$c' \mathbb{1}_{\{|x| \leq c''\}} \leq \rho(x) \leq C e^{-c|x|}$$

where $c', c'', c, C > 0$ are universal constants.

Exercise 3. Prove this proposition.
Hint: for the upper bound, if $\rho(b) < \rho(a)/2$ for some $a < b$, then ρ decays exponentially and in fact $\rho(x) \leq \rho(b) 2^{-x/(b-a)}$ for all $x > b$. As for the lower bound, it's enough to show that $\rho(x) > c'$ for some $x > c''$ and for some $x < -c''$.

Corollary 5 ("reverse Hölder inequalities"). *For any isotropic, log-concave, real-valued random variable X and any $p > -1$,*

$$c \cdot \min\{p+1, 1\} \leq \|X\|_p = (\mathbb{E}|X|^p)^{1/p} \leq C(|p|+1), \tag{6}$$

where $c, C > 0$ are universal constants.

The case $p = 0$ in (6) is interpreted by continuity, i.e.,

$$\|X\|_0 = \exp(\mathbb{E} \log |X|).$$

This is not a norm, yet a nice feature is its multiplicativity: for any random variables X and Y, possibly dependent,

$$\|XY\|_0 = \|X\|_0 \|Y\|_0.$$

Proof of Corollary 5. Begin with the inequality on the right-hand side. By the monotonicity of $p \mapsto \|X\|_p$, it is enough to look at $p > 0$. In this case,

$$\|X\|_p^p = \int_{-\infty}^{\infty} |t|^p \rho(t) dt \leq C \int_{-\infty}^{\infty} |t|^p e^{-c|t|} dt = \frac{2C}{c^{p+1}} \Gamma(p+1) \leq (\widetilde{C} p)^p.$$

where we used the fact that for integer p, we have $\Gamma(p+1) = p! \leq p^p$. For the lower bound, by monotonicity it suffices to look at $p < 0$. Setting $q = -p \in (0,1)$ we have

$$\mathbb{E} \frac{1}{|X|^q} \leq C \int_{-\infty}^{\infty} \frac{1}{|t|^q} e^{-c|t|} dt \leq \frac{C'}{1-q}$$

and hence

$$\|X\|_p = \left(\mathbb{E} \frac{1}{|X|^q} \right)^{-1/q} \geq (C'(1-q))^{1/q} \geq \widetilde{C}(1-q). \qquad \square$$

We proceed to discuss the isoperimetric profile of a log-concave distribution in one dimension. Bobkov [14] shows that for a probability density ρ on the real line,

$$\rho \text{ is log-concave} \qquad \Longleftrightarrow \qquad \rho \circ \Phi^{-1} : [0,1] \to (0, \infty) \text{ is concave} \tag{7}$$

where $\Phi(x) = \int_{-\infty}^{x} \rho(t) dt$ and $\Phi^{-1}(y) = \inf\{x \in \mathbb{R}\,;\, \Phi(x) \geq y\}$. Once stated, (7) is not difficult to prove. It follows from (7) that the function

$$I(x) = \min\left\{\rho \circ \Phi^{-1}, \rho \circ (1-\Phi)^{-1}\right\}$$

is concave. Write μ for the measure whose density is ρ, and note that

$$I(x) = \min\{\rho(\partial H)\,;\, H \text{ is a ray with } \mu(H) = x\}$$

Since the boundary ∂H is a singleton as H is a ray, in this case we abbreviate $\rho(\partial H) = \rho(a)$ if $\partial H = \{a\}$. The following Proposition by Bobkov implies that the concave function I is the *isoperimetric profile* of the probability density ρ.

We prefer to discuss isoperimetry through ε-neighborhoods. For $\varepsilon > 0$ and a subset $A \subseteq \mathbb{R}$ we write $A_\varepsilon = \{x \in \mathbb{R}\,;\, \inf_{y \in A} |x-y| < \varepsilon\}$ for its ε-neighborhood. We remark that analogously to (7), the log-concavity of ρ implies that the function $x \mapsto \Phi(\Phi^{-1}(x) + \varepsilon)$ is concave. This shows that the function

$$I_\varepsilon(x) = \min\{\mu(H_\varepsilon)\,;\, H \text{ is a ray with } \mu(H) = x\}$$

is a concave function of $x \in [0,1]$.

Proposition 6 (Bobkov [14]). *Let μ be a log-concave probability measure on $\mathbb{R}$ with density ρ. Fix $0 < p < 1, \varepsilon > 0$. Then among all Borel subsets $A \subseteq \mathbb{R}$ with $\mu(A) = p$, the infimum of $\mu(A_\varepsilon)$ is attained for a half line.*

Sketch of Proof. It suffices to show that half lines are better than finite unions of intervals. How can we deal with a subset A that is a finite union of intervals? Using the following claim. For $a \in \mathbb{R}$ with $\mu([a,\infty)) > p$ consider the unique interval $J(a) = (a,b)$ such that $\mu(J(a)) = p$. The claim is that the function

$$a \mapsto \mu(J(a)_\varepsilon)$$

is unimodal, thanks to log-concavity (i.e., the function is increasing and then decreasing). Again, once stated this is not too difficult to prove. Given this claim, one may fix all intervals in A but one, and then move the remaining one around and expand and shrink it so as to preserve the total μ-measure. It follows that gluing this interval to one of the sides cannot increase the μ-measure of the ε-neighborhood. $\square$

Combining this with Proposition 4 one gets the following Cheeger type isoperimetry for 1D log-concave measures.

Corollary 7. *Let μ be an isotropic, log-concave probability measure on $\mathbb{R}$ and let $\varepsilon, p \in (0,1)$. Then for any Borel set $S \subseteq \mathbb{R}$ with $\mu(S) = p$,*

$$\mu(S_\varepsilon \setminus S) \geq c \cdot \varepsilon \cdot \min\{p, 1-p\}$$

where $c > 0$ is a universal constant.

Exercise 4. Fill in the details in the proofs of Proposition 6 and Corollary 7.

2 Related functional inequalities

2.1 Cheeger's inequality

Let μ be a probability measure on $\mathbb{R}^n$, or more generally on some metric space (X, d) equipped with its Borel σ-field. The isoperimetric problem for μ asks the following questions: Among sets of given measure, which sets have minimal perimeter? There are several possible notions of perimeter. For our purposes, the most convenient one is the exterior Minkowski content, defined as follows: for every measurable subset A of the ambient space we let

$$\mu_+(A) = \liminf_{\varepsilon \to 0} \frac{\mu(A_\varepsilon \backslash A)}{\varepsilon}.$$

where A_ε is the ε-neighborhood of A, namely the set of points whose distance to A is at most ε. Proposition 6, at the end of the previous section, shows in particular that for 1D log-concave measures, half-lines solve the isoperimetric problem. In higher dimension though, the exact answer to the isoperimetric problem is only known in a handful of very specific cases. For instance, for the Haar measure on the sphere equipped with the geodesic distance, spherical caps (i.e. geodesic balls) are the solution. This is usually attributed to P. Lévy (1922). The answer is also known on Gauss space, and this time affine half-spaces solve the isoperimetric problem. This was proved in 1975 by Sudakov and Tsirelson [79], and independently by Borell [22]. In general solving exactly the isoperimetric problem is hopeless and we content ourselves with a more modest task, such as finding lower bounds on the perimeter of a set A in terms of its measure. When this lower bound is linear, we say that μ satisfies Cheeger's inequality.

Definition 8. *We say that μ satisfies Cheeger's inequality if there is a constant C such that*

$$\min(\mu(A), 1 - \mu(A)) \leq C\mu_+(A), \tag{8}$$

for every measurable set A. The smallest C such that this holds true is called the Cheeger constant, and we denote it ψ_μ below.

For instance, Corollary 7 from the previous section shows that the Cheeger constant of an isotropic log-concave measure in 1D is bounded above by a universal constant.

Remark 9. It is more common to put the constant in the left-hand side of the inequality (8) rather than in the right-hand side. So our Cheeger constant is the reciprocal of the *usual* Cheeger constant.

Cheeger's inequality can be seen as an L^1-Poincaré inequality.

Lemma 10. *Inequality (8) is equivalent to the following:*

$$\min_{c \in \mathbb{R}} \int_X |f - c| \, d\mu \leq C \int_X |\nabla f| \, d\mu, \tag{9}$$

for every Lipschitz function f.

Remark 11. In the right-hand side the quantity $|\nabla f(x)|$ should be interpreted as the local Lipschitz constant of f, namely

$$|\nabla f(x)| = \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x,y)}.$$

This only make sense in a metric space with no isolated points. Actually we will only investigate the case $X = \mathbb{R}^n$ equipped with its usual Euclidean metric from now on.

Remark 12. It is well known that the infimum in the left-hand side is attained at any median for f, i.e. any real c such that both $\mu(f \leq c)$ and $\mu(f \geq c)$ are at least $1/2$.

Proof. We only give a proof sketch, and refer to Bobkov and Houdré [19] (for instance) for more details. The derivation of (9) from (8) relies on the co-area formula: for any Lipschitz f we have

$$\int_X |\nabla f| \, d\mu \geq \int_{\mathbb{R}} \mu^+(f > t) \, dt.$$

In most cases this inequality is actually an equality, but we only need this inequality, which admits a soft proof, again see [19]. Applying Cheeger's inequality to the right-hand side then yields (9). For the converse implication, given a set A, we apply (9) to some suitable Lipschitz approximation of the indicator function of A. A bit more precisely, we pick $\varepsilon_n \to 0$ such that

$$\lim \frac{\mu(A_{\varepsilon_n} \backslash A)}{\varepsilon_n} \to \mu_+(A),$$

we pick another positive sequence (δ_n) tending to 0 (for instance $\delta_n = 1/n$) and we observe that the sequence (f_n) given by

$$f_n = \left(1 - \frac{1}{(1-\delta_n)\varepsilon_n} \cdot d(x, A_{\delta_n \varepsilon_n}) \right)_+$$

satifies $0 \leq f_n \leq 1$ for every n, $f_n \to \mathbb{1}_{\overline{A}}$ pointwise, and $\limsup \int |\nabla f_n| \, d\mu \leq \mu_+(A)$. Applying (9) to f_n and letting n tend to $+\infty$ yields (8) after some computation. ☐

From this version of Cheeger's inequality it is relatively straightfoward to see that Cheeger's inequality is stronger than the Poincaré inequality. Recall from Section 1 that we say that μ satisfies Poincaré if there is a constant C such that

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

for every Lipschitz function f. Also we let $C_P(\mu)$ be the best constant C such that this holds true.

Proposition 13 (Cheeger 1970)**.** *Let μ be a probability measure on $\mathbb{R}^n$ satisfying the Cheeger inequality. Then μ satisfies Poincaré, and we have*

$$C_P(\mu) \leq 4\psi_\mu^2.$$

Remark 14. Maybe it is unfortunate but our convention for the Cheeger constant and Poincaré constant do not have the same homogeneity. The Cheeger constant of a probability measure on $\mathbb{R}^n$ is 1-homogeneous, if we scale μ by a factor λ then the Cheeger constant is multiplied by λ. One the other hand the Poincaré constant is 2-homogeneous.

Proof. Assume that f is Lipschitz and bounded, and has its median at 0. Applying (9) to f_+^2 we get

$$\int_{\mathbb{R}^n} f_+^2 \, d\mu \leq \psi_\mu \int_{\mathbb{R}^n} |\nabla f_+^2| \, d\mu = 2\psi_\mu \int_{\mathbb{R}^n} f_+ |\nabla f_+| \, d\mu.$$

The Cauchy-Schwarz inequality then yields

$$\int_{\mathbb{R}^n} f_+^2 \, d\mu \leq 4\psi_\mu^2 \int_{\mathbb{R}^n} |\nabla f_+|^2 \, d\mu = 4\psi_\mu^2 \int_{\mathbb{R}^n} |\nabla f|^2 \mathbb{1}_{\{f>0\}} \, d\mu.$$

We can do the same with f_- and adding up the two inequalities yields the result. □

The converse inequality is not true in general, one can cook up examples on the line. However it turns out that if we restrict to log-concave measures then the converse is true. This is a result of Buser [26] from 1982, to which we will come back later on in this section.

2.2 Semigroup tools

Let μ be a probability measure on $\mathbb{R}^n$. We do not need log-concavity for now but let us assume that μ is supported on the whole space and has a smooth density ρ. Letting $V = -\log \rho$ be the potential of μ, the Laplace operator associated to μ is the differential operator given by

$$L_\mu = \Delta - \nabla V \cdot \nabla,$$

initially defined on the space of compactly supported smooth functions. For such functions, an integration by parts gives

$$\int_{\mathbb{R}^n} (L_\mu f) g \, d\mu = - \int_{\mathbb{R}^n} \nabla f \cdot \nabla g \, d\mu.$$

This shows in particular that L_μ is symmetric and that $-L_\mu$ is a monotone (unbounded) operator on $L^2(\mu)$. Moreover this operator is known to be essentially self-adjoint, in the sense that its minimal extension is self-adjoint. By a slight abuse of notation we still call L_μ this extension. A bit more explicitly, we call $\mathcal{D}$ the space of functions $f \in L^2(\mu)$ for which there exists a sequence (f_n) of smooth compactly supported functions such that $f_n \to f$ and $(L_\mu f_n)$ converges. The limit of $L_\mu f_n$ does not depend on the choice of the converging sequence (f_n) (this is an immediate consequence of the symmetry of L_μ) and we set $L_\mu f = \lim L_\mu f_n$. The fact that this new L_μ is self adjoint is not quite immediate, not every monotone operator is essentially self adjoint. This has to do with elliptic regularity, we refer to [6, Corollary 3.2.2] for the details. From the integration by parts above we can see that if (f_n) and $(L_\mu f_n)$ converge then also ∇f_n converges. This means that the domain $\mathcal{D}$ contains $H^1(\mu)$ and that the integration by parts $\langle L_\mu f, g \rangle = -\langle \nabla f, \nabla g \rangle$

remains valid for every f, g in the domain. Here the inner product is the one from $L^2(\mu)$, and when we apply it to tensors it has to be interpreted coordinate wise. Being self-adjoint and monotone (negative) the operator L_μ admits a spectral decomposition

$$L_\mu = -\int_0^\infty \lambda \, dE_\lambda. \tag{10}$$

The semigroup associated to L_μ is then defined as

$$P_t = \mathrm{e}^{tL_\mu} = \int_0^\infty \mathrm{e}^{-t\lambda} \, dE_\lambda.$$

For fixed t the operator P_t is a self-adjoint bounded operator in $L^2(\mu)$ and we have the semigroup property $P_t \circ P_s = P_{t+s}$. If f is a fixed function of $L^2(\mu)$ the function $F(t,x) = P_t f(x)$ is the solution to the parabolic equation

$$\begin{cases} F(0,\cdot) = f \\ \partial_t F = L_\mu F, \end{cases}$$

at least in a weak sense.

We now move on to the probabilistic representation of the semigroup (P_t). Consider the diffusion (X_t) given by

$$dX_t = \sqrt{2} \cdot dW_t - \nabla V(X_t) \, dt, \tag{11}$$

where (W_t) is standard Brownian motion. Then (X_t) is a Markov process, and (P_t) is the corresponding semigroup. Namely for every test function f we have

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

where the subscript x next to the expectation denotes the starting point of (X_t). This allows to prove inequalities for the semigroup (P_t) using probabilistic techniques. The next result is considered folklore, see e.g. [6, section 9.9.] for some historical perspectives.

Lemma 15. *If μ is log-concave then Lipschitz functions are preserved along the semigroup, and moreover $\|P_t f\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$ for every f and every $t > 0$.*

Proof. Let $x, y \in \mathbb{R}^n$, and let (X_t^x) and (X_t^y) be two solutions of the SDE (11) using the same Brownian motion, but starting at two different points x and y. This is called parallel coupling. Then the process $(X_t^x - Y_t^x)$ is an absolutely continuous function of t (the Brownian part cancels out). Moreover, thanks to the convexity of V,

$$\frac{d}{dt}|X_t^x - X_t^y|^2 = -2(X_t^x - X_t^y) \cdot (\nabla V(X_t^x) - \nabla V(X_t^y)) \leq 0.$$

So the distance $|X_t^x - X_t^y|$ is almost surely decreasing. Therefore its expectation is also decreasing, and in particular

$$\mathbb{E}|X_t^x - X_t^y| \leq |x - y|.$$

Now suppose f is a Lipschitz function. Then from the previous inequality we get

$$|P_t f(x) - P_t f(y)| = |\mathbb{E} f(X_t^x) - \mathbb{E} f(X_t^y)| \leq \mathbb{E}|f(X_t^x) - f(X_t^y)| \leq \|f\|_{\mathrm{Lip}} \cdot |x - y|,$$

which is the result. □

The next result seems to be due to Varopoulos [82].

Proposition 16. *Suppose μ is log-concave. Then for every bounded function f and every $t > 0$ the function $P_t f$ is Lipschitz and moreover*

$$\|P_t f\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{t}} \cdot \|f\|_\infty.$$

Proof. Again we use a coupling argument, see [60] for an alternate argument using only analytic tools. Suppose that f is a bounded function. Fix $x, y \in \mathbb{R}^n$, and let (X_t^x) and (X_t^y) be two processes solving the SDE (11) initiated at x and y respectively. Then

$$|P_t f(x) - P_t f(y)| \leq \mathbb{E}|f(X_t^x) - f(X_t^y)| \leq 2\|f\|_\infty \cdot \mathbb{P}(X_t^x \neq X_t^y). \tag{12}$$

It remains to choose a coupling for which the right-hand side is small. Parallel coupling is awful here, as it actually prevents X_t^x and X_t^y from meeting. Instead, we choose the Brownian increment for X_t^y to be the reflection of that of X_t^x with respect to the hyperplane $(X_t^x - X_t^y)^\perp$. If (W_t) is the Browian motion for X_t^x, the equation for X_t^y is thus

$$dX_t^y = \sqrt{2} \cdot \left(\mathrm{Id} - 2v_t^{\otimes 2}\right) dW_t - \nabla V(X_t^y)\, dt$$

where (v_t) is the unit vector $(X_t^x - X_t^y)/|X_t^x - X_t^y|$. Actually we do so until the first time (denoted τ) when the two processes meet. After time τ we just set $X_t^y = X_t^x$. We will not justify properly here why this is well defined, but this coupling technique, usually referred to as *mirror coupling*, is a relatively standard tool, see for instance [65]. Itô's formula shows that up to the coupling time τ the equation for the distance between the two processes is

$$d|X_t^x - X_t^y| = 2\sqrt{2} v_t \cdot dW_t - v_t \cdot (\nabla V(X_t^x) - \nabla V(X_t^y))\, dt.$$

Itô's term vanishes because the Brownian increment takes place in a direction where the Hessian matrix of the norm vanishes. Once again, in the log-concave case the second term from the right hand side is negative. Notice also that $B_t := \int_0^t v_s \cdot dW_s$ is a standard (one dimensional) Brownian motion. Therefore up to the coupling time τ we have

$$|X_t^x - X_t^y| \leq |x - y| + 2\sqrt{2} B_t,$$

where (B_t) is some standard one dimensional Brownian motion. Therefore

$$\mathbb{P}(X_t^x \neq X_t^y) = \mathbb{P}(\tau > t) \leq \mathbb{P}\left(\forall s \leq t\colon\, B_s > -\frac{|x-y|}{2\sqrt{2}}\right).$$

By the reflection principle for the Brownian motion

$$\mathbb{P}\left(\exists s \leq t\colon\, B_s \leq -\frac{|x-y|}{2\sqrt{2}}\right) = 2 \cdot \mathbb{P}\left(B_t \leq -\frac{|x-y|}{2\sqrt{2}}\right) = \mathbb{P}\left(|g| \geq \frac{|x-y|}{2\sqrt{2t}}\right)$$

where g is a standard Gaussian variable. Hence the inequality

$$\mathbb{P}(X_t^x \neq X_t^y) \leq \Psi\left(\frac{|x-y|}{2\sqrt{2t}}\right),$$

where $\Psi(r) = (2/\pi)^{1/2} \int_0^r \mathrm{e}^{-u^2/2}\, du$ is the distribution function of $|g|$. Recalling (12) and taking the supremum over x, y gives

$$\|P_t f\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{2t}} \cdot \sup_{a>0} \left\{ \frac{\Psi(a)}{a} \right\} \cdot \|f\|_\infty.$$

The expression inside the sup is decreasing, so the sup equals the limit as a tends to 0, which is $(2/\pi)^{1/2}$. We thus get the desired inequality (even with a better constant than announced). □

The next corollary is taken from Ledoux [58].

Corollary 17. *If μ is log-concave, then for every locally Lipschitz function f we have*

$$\|f - P_t f\|_{L^1(\mu)} \leq 2\sqrt{t} \cdot \||\nabla f|\|_{L^1(\mu)}.$$

Also for every measurable set A we have

$$\mu(A)(1-\mu(A)) = \mathrm{Var}_\mu(\mathbb{1}_A) \leq \sqrt{2t} \cdot \mu^+(A) + \mathrm{Var}_\mu(P_t \mathbb{1}_A).$$

Proof. Let f be a Lipschitz function and g be a smooth bounded function. Using the fact that the semigroup is self adjoint, and the integration by part formula, we get

$$\langle f - P_t f, g \rangle = \langle f, g - P_t g \rangle = -\int_0^t \langle f, L P_s g \rangle\, dt = \int_0^t \langle \nabla f, \nabla P_s g \rangle\, ds.$$

By the previous proposition,

$$\langle \nabla f, \nabla P_s g \rangle \leq \||\nabla f|\|_{L^1(\mu)} \cdot \|P_s g\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{s}} \||\nabla f|\|_{L^1(\mu)} \|g\|_\infty.$$

Integrating between 0 and t and plugging back in the previous display we get

$$\langle f - P_t f, g \rangle \leq 2\sqrt{t} \cdot \||\nabla f|\|_{L^1(\mu)} \|g\|_\infty,$$

which is the result. For the second inequality, applying the first one to a suitable Lipschitz approximation of the indicator function of A, as in the proof of Lemma 10, we get

$$\|\mathbb{1}_A - P_t \mathbb{1}_A\|_1 \leq 2\sqrt{t} \cdot \mu^+(A).$$

Moreover, using reversibility, it is not hard to see that

$$\|\mathbb{1}_A - P_t \mathbb{1}_A\|_1 = 2\left(\mathrm{Var}_\mu(\mathbb{1}_A) - \mathrm{Var}_\mu(P_{t/2} \mathbb{1}_A)\right).$$

Hence the result. □

2.3 A result of E. Milman

We said earlier that the inequality $C_P(\mu) \leq C\psi_\mu^2$ can be reversed in the log-concave case. Actually we will prove a much stronger statement, which is due to E. Milman.

Definition 18. *If μ is a probability measure on $\mathbb{R}^n$, the function*

$$I_\mu \colon r \in [0,1] \mapsto \inf\{\mu_+(S) \colon \mu(S) = r\}.$$

is called the isoperimetric profile of μ.

With this definition Cheeger's inequality can be rewritten as

$$\psi_\mu \cdot I_\mu(r) \geq \min(r, 1-r).$$

The following is a deep result from geometric measure theory.

Theorem 19. *The isoperimetric profile of a log-concave measure is concave.*

We will use this as a blackbox, we refer to the appendix of [67] for an historical account and the relevant references. Another good reference for this is Bayle's Ph.D. thesis [11] (if you read french). This has important implications for us. Indeed, since the isoperimetric profile is non negative, its concavity implies that

$$I_\mu(t) \geq 2 \cdot I_\mu(1/2) \min(t, 1-t).$$

In particular the Cheeger constant of μ satisfies

$$\psi_\mu \leq \frac{1}{2 \cdot I_\mu(1/2)}. \tag{13}$$

Therefore, for a log-concave measure, in order to prove Cheeger's inequality, it is enough to look at the perimeter of sets of measure $1/2$. Combining this information with the results from the previous section we arrive at the following.

Theorem 20. *If μ is log-concave, then there exists a 1-Lischitz function f satisfying*

$$\|f\|_\infty^2 \approx \mathrm{Var}_\mu(f) \approx \psi_\mu^2.$$

Here the symbol $\approx$ means that the ratio between the two quantities is comprised between two positive universal constants. Theorem 20 is essentially due to E. Milman [67]. The proof we give is very much inspired by Ledoux's proof of Buser's inequality [58].

Proof. By (13) if A is a set of measure $1/2$ that has near minimal surface, say up to a factor 2, then

$$\mu_+(A) \leq \frac{1}{\psi_\mu}. \tag{14}$$

Let $t > 0$. By Corollary 17, and since $\mu(A) = 1/2$,

$$\frac{1}{4} \leq \sqrt{2t} \cdot \mu_+(A) + \mathrm{Var}_\mu(P_t \mathbb{1}_A) \leq \frac{\sqrt{2t}}{\psi_\mu} + \mathrm{Var}_\mu(P_t \mathbb{1}_A).$$

If t is a sufficiently small multiple of ψ_μ^2 we thus get $\mathrm{Var}_\mu(P_t \mathbb{1}_A) \geq \frac{1}{8}$ (say). On the other hand, by Proposition 16,

$$\|P_t \mathbb{1}_A\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{t}} \leq \frac{C}{\psi_\mu},$$

for some constant C. Putting everything together we see that the function $f = (\psi_\mu / C) \cdot P_t \mathbb{1}_A$ is 1-Lipschitz and satisfies

$$\psi_\mu^2 \lesssim \mathrm{Var}_\mu(f) \leq \|f\|_\infty^2 \lesssim \psi_\mu^2 \qquad \square.$$

Note that since f is 1-Lipschitz, the Poincaré inequality yields $\mathrm{Var}_\mu(f) \leq C_P(\mu)$. The result above thus implies that

$$\psi_\mu^2 \lesssim C_P(\mu).$$

In other words, the Cheeger inequality can be reversed in the log-concave case. Moreover, the theorem actually yields a lot more. It implies that it is enough to bound the variance of Lipschitz functions to get Poincaré (or Cheeger). More precisely, we get the following.

Corollary 21 (Buser [26], Ledoux [58], E. Milman [67])**.** *For any log-concave measure μ,*

$$\psi_\mu^2 \approx C_P(\mu) \approx \sup\left\{\mathrm{Var}_\mu(f) \colon \|f\|_{\mathrm{Lip}} \leq 1\right\}.$$

Constants are mostly regarded as irrelevant in theses notes but let us mention that for the left-most equality, the optimal constants are actually known. Indeed De Ponti and Mondino [33] proved that

$$\frac{1}{\pi}\psi_\mu^2 \leq C_P(\mu) \leq 4\psi_\mu^2.$$

In section 3.2 we give another proof of this corollary based on L_1 transportation that avoids the concavity of the isoperimetric profile blackbox.

Let us also point out that the corollary does not quite use the full strength of Theorem 20, it does not use the information about the L^∞ norm of f. So we actually have stronger form of the corollary. Namely, in the log-concave case, to get Cheeger, or Poincaré, it is enough to bound the variance of a bounded Lipschitz function whose Lipschitz constant is 1, and whose L^∞-norm is of the same order as its standard deviation.

2.4 Concentration of measure

Definition 22. *Let (X, d, μ) be a metric measure space. The concentration function of μ is defined by*

$$\alpha_\mu \colon r \mapsto \sup\left\{1 - \mu(S_r) \colon \mu(S) \geq 1/2\right\}$$

where S_r is the r-neighborhood of the set S.

As for isoperimetry, we can only compute the exact value of the concentration function in some very specific models such as the uniform measure on the sphere or the Gaussian measure. In general we are happy with a good upper bound for α_μ. The most interesting types of upper bounds for us are the case of Gaussian concentration and of exponential concentration.

Definition 23. *We say that μ satisfies Gaussian concentration if there is a constant C_G such that*

$$\alpha_\mu(r) \leq 2 \cdot \exp\left(-\frac{r^2}{C_G}\right), \quad \forall r \geq 0.$$

We say that μ satisfies exponential concentration if there exists a constant C_{exp} such that

$$\alpha_\mu(r) \leq 2 \cdot \exp\left(-\frac{r}{C_{exp}}\right), \quad \forall r \geq 0.$$

Moreover the smallest constants C_G, C_{exp} such that the above inequalities hold true are called the Gaussian concentration constant and the exponential concentration constant, respectively.

We are interested here in concentration properties of log-concave measures on $\mathbb{R}^n$. Gaussian concentration cannot be true in general (think of μ being the exponential measure) but there is no obstruction to having exponential concentration with a dimension free constant for isotropic log-concave measures, and this is in fact equivalent to the KLS conjecture from the previous section. Indeed, it is well-known that the Poincaré inequality yields exponential concentration, and more precisely that for any probability measure μ on $\mathbb{R}^n$ satisfying the Poincaré inequality we have

$$\alpha_\mu(r) \leq 2 \cdot \exp\left(-\frac{r}{L \cdot \sqrt{C_P(\mu)}}\right), \quad \forall r \geq 0,$$

where L is a universal constant. We will skip the derivation of this from Poincaré here, but this is not very hard, see for instance [6, section 4.4.2].

Once again, in the log-concave case this implication can be reversed. Indeed, by E. Milman's theorem (Corollary 21) from the previous subsection the Poincaré constant is a largest variance of a 1-Lipschitz function (up to a constant). If f is 1-Lipshitz, by definition of the concentration function we have

$$\mu(f - m \geq r) \leq \alpha_\mu(r),$$

for every $r > 0$, and where m is a median for f. From this we obtain easily

$$\mathrm{Var}_\mu(f) \leq 4 \int_0^\infty r \cdot \alpha_\mu(r) \, dr.$$

Therefore, in the log-concave case

$$C_P(\mu) \lesssim \int_0^\infty r \cdot \alpha_\mu(r) \, dr. \tag{15}$$

This implies in particular that the Poincaré constant of μ and the exponential concentration constant squared are actually of the same order.

2.5 Log-Sobolev and Talagrand

We have seen earlier that Poincaré is weaker than Cheeger in general but equivalent to it within the class of log-concave measures. We shall see now that log-concavity also allows to reverse the hierarchy between the log-Sobolev inequality and the transportation inequality. A probability measure μ on $\mathbb{R}^n$ is said to satisfy the logarithmic Sobolev inequality if there exists a constant $C > 0$ such that

$$D(\nu \mid \mu) \leq \frac{C}{2} I(\nu \mid \nu)$$

for every probability measure ν, where $D(\nu \mid \mu)$ and $I(\nu \mid \mu)$ denote the relative entropy and Fisher information, respectively:

$$D(\nu \mid \mu) = \int_{\mathbb{R}^n} \log(\frac{d\nu}{d\mu})\, d\nu \quad \text{and} \quad I(\nu \mid \mu) = \int_{\mathbb{R}^n} |\nabla \log(\frac{d\nu}{d\mu})|^2\, d\nu.$$

The best constant C is called the log-Sobolev constant, denoted $C_{LS}(\mu)$ below. The factor $1/2$ is just a matter of convention. With this convention the log-Sobolev constant of the standard Gaussian 1. This is a stronger inequality than Poincaré. More precisely we have $C_P(\mu) \leq C_{LS}(\mu)$ for any μ. This is easily seen by applying log-Sobolev to a probability measure whose density with respect to μ is $1 + \varepsilon f$ and letting ε tend to 0. Not every log-concave measure satisfy log-Sobolev, simply because log-Sobolev implies sub-Gaussian tails, so for instance the exponential measure (on $\mathbb{R}$) does not statisfy log-Sobolev. A bit more precisely, log-Sobolev implies Gaussian concentration: if μ satisfies log-Sobolev then for any set S we have

$$\mu(S)(1 - \mu(S_r)) \leq \exp\left(-c \cdot \frac{r^2}{C_{LS}(\mu)}\right).$$

We will come back to that later on.

Recall that if μ, ν are probability measures on $\mathbb{R}^n$, the quadratic transportation cost from μ to ν is defined as

$$T_2(\nu, \nu) = \inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2\, d\pi \right\},$$

where the infimum is taken over every coupling π of μ and ν, namely every probability measure on the product space whose marginals are μ and ν. In the next section we will speak about the Monge transport cost, which is the L^1 version of this.

Proposition 24 (Otto and Villani [71])**.** *If μ satisfies log-Sobolev then for every probability measure ν we have*

$$T_2(\nu, \mu) \leq 2C_{LS}(\mu) \cdot D(\nu \mid \mu).$$

This transportation/entropy inequality is sometimes called Talagrand's inequality, as it was first established by Talagrand for the Gaussian measure, see [81]. Again in the log-concave case the implication log-Sobolev/Talagrand can be reversed. Indeed, we have the following, also due to Otto and Villani.

Proposition 25 (Otto and Villani [71]). *If μ is log-concave then for every probability measure ν on $\mathbb{R}^n$ we have*

$$D(\nu \mid \mu) \leq \sqrt{T_2(\nu, \mu) \cdot I(\nu \mid \mu)}.$$

This is only a particular case of the Otto-Villani result, there is also a version for semi-log-concave measures, namely measures for which we have a possibly negative lower bound on the Hessian of the potential. This inequality goes by the name HWI. The reason for this name is not apparent from our choice of notations, but relative entropy is often denoted H, and the transport cost T_2 can also be denoted W_2 or rather W_2^2 (for Wasserstein). From the HWI inequality we see that the implication between log-Sobolev and Talagrand can be reversed for log-concave measures: if we happen to know

$$T_2(\nu, \mu) \leq C_2 D(\nu \mid \mu)$$

for μ log-concave, then we get log-Sobolev for μ and $C_{LS}(\mu) \leq 2C_2$. We will not spell out the proofs of the Otto-Villani results here and we refer to [71] (see also [18]).

We have seen above that the equivalence between Cheeger and Poincaré can be considerably reinforced. This is also the case here, and this is yet again a result of E. Milman.

Theorem 26 (E. Milman [68]). *For a log-concave probability measure we have equivalence between Gaussian concentration and the log-Sobolev inequality, and moreover the log-Sobolev constant and the Gaussian concentration constant are within a universal factor of each other.*

Proof. There are several proofs of this result in the literature, see [68, 61]. The proof sketch that we give here is taken from Gozlan, Roberto, Samson [41]. We said earlier that log-Sobolev implies Gaussian concentration, but a bit more is true: the weaker Talagrand inequality also implies Gaussian concentration. Let us explain why briefly. By some convex duality principle, T_2 can be also expressed as a supremum, namely

$$T_2(\mu, \nu) = \sup_f \left\{ \int_{\mathbb{R}^n} Q_{1/2} f \, d\mu - \int_{\mathbb{R}^n} f \, d\nu \right\}$$

where the $Q_t f$ is the infimum convolution of f with some multiple of the distance squared:

$$Q_t f(x) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2t} |x - y|^2 \right\}.$$

It can also be shown that (Q_t) is a semigroup of operators, namely we have $Q_s Q_t = Q_{s+t}$. Lastly there is also some duality between the log-Laplace transform and the relative entropy:

$$\log \int_{\mathbb{R}^n} \mathrm{e}^f \, d\mu = \sup_\nu \left\{ \int_{\mathbb{R}^n} f \, d\nu - D(\nu \mid \mu) \right\},$$

where the supremum is taken over every probability measure ν. Using all this, it is pretty easy to see that Talagrand's inequality

$$T_2(\nu, \mu) \leq 2C_T \cdot D(\nu \mid \mu), \quad \forall \nu$$

is equivalent to

$$\int_{\mathbb{R}^n} \exp(Q_{C_T} f)\, d\mu \leq \exp\left(\int_{\mathbb{R}^n} f\, d\mu\right), \quad \forall f.$$

Applying this to both $Q_{C_T} f$ and $-Q_{C_T} f$, using the fact that (Q_t) is a semigroup, and multiplying the two inequalities together we get

$$\int_{\mathbb{R}^n} \exp(Q_{C_T}(-Q_{C_T} f)\, d\mu \cdot \int_{\mathbb{R}^n} \exp(Q_{2C_T} f)\, d\mu \leq 1.$$

But clearly $-f \leq Q_{C_T}(-Q_{C_T} f)$, so we obtain

$$\int_{\mathbb{R}^n} \exp(-f)\, d\mu \cdot \int_{\mathbb{R}^n} \exp(Q_{2C_T} f)\, d\mu \leq 1.$$

Applying to $f = -\log \mathbb{1}_A$ we get

$$\int_{\mathbb{R}^n} \exp\left(\frac{d(x,A)^2}{2C_T}\right)\, dx \leq \frac{1}{\mu(A)},$$

for every set A. By Markov inequality this implies

$$\alpha_\mu(r) \leq 2 \cdot \exp\left(-\frac{r^2}{2C_T}\right).$$

So Talagrand implies Gaussian concentration, and moreover the Gaussian concentration constant is at most the constant in Talagrand, up to a factor 2. Now we want to reverse this, so we assume

$$\alpha_\mu(r) \leq 2\mathrm{e}^{-r^2/C_G}.$$

It is easily seen to imply

$$\int_{\mathbb{R}^n} \exp(Q_{2C_G} f)\, d\mu \lesssim \exp(m_f).$$

for every f, and where m_f is a median for f. Again the notation $\lesssim$ means up to a universal factor. Again, applying this $-Q_{2C_G} f$ and $Q_{2C_G} f$ and multiplying the two inequalities together we get

$$\int_{\mathbb{R}^n} \mathrm{e}^{-f}\, d\mu \cdot \int_{\mathbb{R}^n} \exp(Q_{4C_G} f)\, d\mu \lesssim 1,$$

hence by Jensen's inequality

$$\int_{\mathbb{R}^n} \exp(Q_{4C_G} f)\, d\mu \lesssim \exp\left(\int_{\mathbb{R}^n} f\, d\mu\right).$$

In other words we get the dual version of Talagrand, but with some prefactor. In terms of transport and entropy this gives

$$T_2(\nu,\mu) \lesssim C_G(D(\nu \mid \mu) + 1).$$

So we have an additional additive constant in the right-hand side of Talagrand. We have not used log-concavity yet, this would be true for any measure satisfying Gaussian concentration. Now assuming log-concavity, we can plug this into the HWI inequality (Proposition 25). We get

$$D(\nu \mid \mu) \lesssim C_G \cdot I(\nu \mid \mu) + 1.$$

Again, we get some weak form of log-Sobolev with an additional constant term in the right-hand side. This is sometimes called non-tight log-Sobolev inequality. To get rid of that constant, observe first that we clearly have from the first theorem of E. Milman (see equation (15))

$$C_P(\mu) \lesssim C_G.$$

Moreover, non-tight log-Sobolev can be reformulated as

$$\mathrm{ent}_\mu(f^2) \lesssim C_G \int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu + \int_{\mathbb{R}^n} f^2 \, d\mu,$$

where the entropy of a non negative function f is defined as

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

Now there is a nice inequality by Rothaus [77] which states that for any $f \colon \mathbb{R}^n \to \mathbb{R}$ and any constant c we have

$$\mathrm{ent}_\mu((f+c)^2) \leq \mathrm{ent}_\mu(f^2) + 2 \int_{\mathbb{R}^n} f^2 \, d\mu.$$

Using this inequality it is easy to see that our non tight version of log-Sobolev and the bound that we have on $C_P(\mu)$ altogether imply

$$\mathrm{ent}_\mu(f) \lesssim C_G \int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu,$$

which is a reformulation of the desired log-Sobolev inequality. □

3 Optimal transport theory with the Monge cost

Let μ_1 and μ_2 be two measures in $\mathbb{R}^n$, say compactly-supported and absolutely continuous, with the same total mass, i.e., $\mu_1(\mathbb{R}^n) = \mu_2(\mathbb{R}^n)$. We would like to push-forward the measure μ_1 to the measure μ_2 in the most efficient way, that minimizes the average distance that points have to travel. That is, we look at the optimization problem

$$\inf_{S_*(\mu)=\nu} \int_{\mathbb{R}^n} |Sx - x| \, \mu_1(dx).$$

This is the problem of Optimal Transport with the Monge cost or the L^1 cost, considered by Monge in 1781. See Cayley's review of Monge's work [29] from 1883. For a more recent survey on Monge's problem, see for instance [21]. Here is a heuristics from Monge's paper that explains why this problem induces a partition into segments.

Monge heuristic: For the optimal transport map T, the segments $(x, T(x))$ $(x \in Supp(\mu_1))$ do not intersect, unless they overlap.

Explanation. Suppose that open segments (x, Tx) and (y, Ty) are not parallel and intersect at a point z. The triangle inequality then shows that $|x - Ty| + |y - Tx| < |x - Tx| + |y - Ty|$, which contradicts the fact that the map T is optimal. □

This is related to the following elementary riddle: given 50 red points and 50 blue points in the plane, in general position, find a matching so that the corresponding segments do not intersect.

Since the above argument relies only on the triangle inequality, you would expect that the optimal transport problem would induce a partition into geodesics also for Riemannian manifolds, or Finslerian manifolds, or measure metric spaces of some type – basically wherever the triangle inequality holds true (under some regularity assumptions).

3.1 Linear programming relaxation and the dual problem

In Monge's problem we minimize over all maps S that push-forward μ_1 to μ_2. There is a relaxation of this problem, that looks at all possible *couplings*, or transport plans, of the two distributions. That is, instead of mapping a point x to a single point Tx, we are allowed to spread the mass across a region. Thus we look at all measures γ on $\mathbb{R}^n \times \mathbb{R}^n$ with

$$(\pi_1)_*\gamma = \mu_1 \qquad \text{and} \qquad (\pi_2)_*\gamma = \mu_2.$$

where $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$. Such a measure is called a *coupling* of μ and ν. In other words, we now look at *transport plans* rather than *transport maps*. The advantage is that the space of all couplings is a convex set. The relaxed optimal transport problem involves minimizing the average distance that points travel, namely we look at

$$\inf_{(\pi_1)_*\gamma=\mu, (\pi_2)_*\gamma=\nu} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y| \, \gamma(dx, dy).$$

Hence we minimize a linear function on a convex set, this is Linear Programming or Functional Analysis (see e.g. Kantorovich and Akilov [46, Section VIII.4]).

Theorem 27. *(The dual problem) Let μ_1, μ_2 be two absolutely continuous measures in $\mathbb{R}^n$ with the same total mass. Assume that*

$$\int_{\mathbb{R}^n} |x|\,\mu_1(dx) < \infty \qquad \text{and} \qquad \int_{\mathbb{R}^n} |x|\,\mu_2(dx) < \infty.$$

Denote $\mu = \mu_2 - \mu_1$. Then the following quantities are equal:

1. *The minimum over all couplings γ of μ_1 and μ_2 of the integral*

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|\,\gamma(dx,dy).$$

2. *The maximum over all 1-Lipschitz functions $u : \mathbb{R}^n \to \mathbb{R}$ of*

$$\int_{\mathbb{R}^n} u(x)\,\mu(dx)$$

3. *The minimum over all maps T with $T_*\mu_1 = \mu_2$ of*

$$\int_{\mathbb{R}^n} |x - Tx|\,\mu_1(dx).$$

Proof sketch. We refer to Ambrosio [3] for full details. For the easy direction of the linear programming duality, pick a 1-Lipschitz map u and a coupling γ. For any points $x, y \in \mathbb{R}^n$,

$$u(y) - u(x) \leq |x-y|.$$

Integrating with respect to γ, we get

$$\int_{\mathbb{R}^n} u d\mu = \int_{\mathbb{R}^n \times \mathbb{R}^n} [u(y) - u(x)]\,\gamma(dx,dy) \leq \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|\,\gamma(dx,dy). \tag{16}$$

Hence we need to find u and γ so that equality is attained in (16). The argument goes roughly as follows. A compactness argument shows that the infimum over all couplings is attained. Indeed, by Alaoglu's theorem, the collection of all couplings is compact in the w^*-topology (integration against continuous functions on $\mathbb{R}^n$ whose limit at infinity exists). The functional $\gamma \mapsto \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|\,\gamma(dx,dy)$ is lower semi-continuous in w^*-topology, hence its minimum is attained.

Similarly to the Monge heuristics, the optimality implies that the support of γ must be cyclically monotone: If $(x_i, y_i) \in Supp(\gamma) \subseteq \mathbb{R}^n \times \mathbb{R}^n$ for $i = 1, \ldots, N$ then for any permutation $\sigma \in S_N$,

$$\sum_{i=1}^{N} |x_i - y_i| \leq \sum_{i=1}^{N} |x_i - y_{\sigma(i)}|. \tag{17}$$

Indeed, otherwise one may pick small balls around x_i and y_i and rearrange them to contradict optimality. Similarly to Rockafellar's theorem [75] from convex analysis, condition (17) implies that there exists a 1-Lipschitz function $u : \mathbb{R}^n \to \mathbb{R}$ with

$$(x, y) \in Supp(\gamma) \qquad \Longrightarrow \qquad u(y) - u(x) = |y - x|. \tag{18}$$

Indeed, fix $(x_0, y_0) \in Supp(\gamma)$ and define $u(x)$ as the supremum over all lower bounds with $u(x_0) = 0$,

$$u(x) = \sup_{N,(x_1,y_1),\dots,(x_N,y_N)\in Supp(\gamma)} \{|x_0 - y_0| - |y_0 - x_1| + |x_1 - y_1| - |y_1 - x_2| + \dots - |y_N - x|\}$$

It follows from (17) that $u(x_0) = 0$. The function u is a 1-Lipschitz function as a supremum of 1-Lipschitz functions. It follows from the definition of u that (18) holds true. Hence we found u and γ so that equality is attained in (16). The proof that γ can also be replaced by a transport map is due to Evans and Gangbo [37]. This relies on analysis of the structure of u that will be described next. □

Remark 28. The minimizers γ or T are not at all unique. It is actually the 1-Lipschitz function u which is essentially determined. More precisely, the gradient ∇u is determined μ-almost everywhere.

We move on to discuss the structure of 1-Lipschitz functions. Observe that when a 1-Lipschitz function u satisfies $|u(x) - u(y)| = |x - y|$, for some points $x, y \in \mathbb{R}^n$, it necessarily grows in speed one along the segment from x to y. A maximal open segment I on which u grows with speed one, i.e., $|u(x) - u(y)| = |x - y|$ for all $x, y \in I$, is called a *transport ray*. Theorem 27 tells us that optimal transport only happens only along transport rays, we only rearrange mass along transport rays.

It is illuminating to draw the transport rays of the function $u(x) = x_1$ in connection with Fubini's theorem

$$\int_{\mathbb{R}^2} \varphi = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \varphi(x_1, x_2) dx_1 \right) dx_2,$$

and of the function $u(x) = |x|$ on $\mathbb{R}^2 \cong \mathbb{C}$ in connection with integration in polar coordinates:

$$\int_{\mathbb{R}^2} \varphi = \int_0^{2\pi} \left(\int_0^{\infty} \varphi(re^{i\theta}) r dr \right) d\theta.$$

Note that the Jacobian factor on the needle is log-concave in both examples.

The next step is to understand the *disintegration of measure* or *conditional probabilities* induced by the partition into transport rays. Let u be a maximizer as above, with

$$\mu = \mu_2 - \mu_1,$$

and with the two measures satisfying the requirements of Theorem 27. As it turns out, it is guaranteed that transport rays of positive length form a partition of the entire support of the measure μ, up to a set of measure zero. Write

$$f = \frac{d\mu}{d\lambda}$$

where λ is any log-concave reference measure in $\mathbb{R}^n$ (not necessarily finite; it could be the Lebesgue measure for instance). The assumption that $\mu_1(\mathbb{R}^n) = \mu_2(\mathbb{R}^n)$ is equivalent to the requirement that

$$\int_{\mathbb{R}^n} f d\lambda = 0. \tag{19}$$

The following theorem requires careful regularity analysis, and in addition to Evans and Gangbo [37] it builds upon works by Caffarelli, Feldman and McCann [27] as well as [50]. It is analogous to integration in polar coordinates, yet with respect to a general 1-Lipschitz guiding function, rather than just $u(x) = |x|$. In the following theorem a line segment could also mean a singleton, a ray or a line.

Theorem 29 (Evans and Gangbo [37], Caffarelli, Feldman and McCann [27], Klartag [50])**.** *Let λ be an absolutely-continuous, log-concave measure on $\mathbb{R}^n$, and let $f \in L^1(\lambda)$ satisfy (19). Then there is a collection Ω of line segments that form a partition of $\mathbb{R}^n$, a family of measures $\{\lambda_\mathcal{I}\}_{\mathcal{I}\in\Omega}$, and a measure ν on the space of segments Ω, such that*

1. *For any $\mathcal{I} \in \Omega$ the measure $\lambda_\mathcal{I}$ is supported on the line segment $\mathcal{I}$. If $\mathcal{I}$ is of non-zero length, then it is a transport ray of the 1-Lipschitz function u.*
2. *Disintegration of measure*
$$\lambda = \int_\Omega \lambda_\mathcal{I}\, \nu(d\mathcal{I}).$$
3. *Mass balance condition: for ν-almost any $\mathcal{I} \in \Omega$,*
$$\int_\mathcal{I} f d\lambda_\mathcal{I} = 0.$$
4. *For ν-almost any $\mathcal{I} \in \Omega$, the measure $\lambda_\mathcal{I}$ has a C^∞-smooth, positive density ρ with respect to the Lebesgue measure on the segment $\mathcal{I}$ which is log-concave.*
(In fact, in the case where λ is the Lebesgue measure, it is a polynomial of degree $n-1$ with real roots, that does not vanish in the support of $\lambda_\mathcal{I}$).

Remark 30. This theorem may be generalized to any Riemannian manifold with non-negative Ricci curvature. We replace the line segment $\mathcal{I}$ by a unit-speed geodesic $\gamma = \gamma_\mathcal{I}$, and set $\kappa(t) = Ricci(\dot\gamma(t), \dot\gamma(t)), n = \dim(M)$. Denote by $\rho = \rho_\mathcal{I}$ the density of $\mu_\mathcal{I}$ with respect to arclength on the geodesic $\gamma = \gamma_\mathcal{I}$. Then,

$$\left(\rho^{\frac{1}{n-1}}\right)'' + \frac{\kappa}{n-1} \cdot \rho^{\frac{1}{n-1}} \leq 0.$$

The Riemannian version may be used to prove isoperimetric inequalities under lower bounds on the Ricci curvature, as well as Poincaré inequalities, log-Sobolev inequalities, Brunn-Minkowski inequalities and more, see [50]. A generalization to the context of synthetic Ricci bounds was introduced by Cavalleti and Mondino [28]. See also Ohta [70] for the non-symmetric, Finslerian case.

Some ideas from the proof of Theorem 29. The proof of Theorem 29 does not use sophisticated results from Geometric Measure Theory, but it consists of several steps. Essentially,

- Show that a 1-Lipschitz u is always differentiable in the relative interior of a transport ray.
- The next step is to show that ∇u is a locally-Lipschitz function on a set which is only slightly smaller than the union of all transport rays, and that the restriction of u to this set may be extended to a $C^{1,1}$-function on $\mathbb{R}^n$.
- This is just enough regularity in order to allow change of variables in an integral, which yields the disintegration.
- By differentiating the Jacobian one sees that the logarithmic derivative of the needle density is the mean curvature of the level set of u, and the inverse principal curvatures grow linearly along the needle. This yields log-concavity along each needle.
- The mass balance condition follows from the fact that γ is a coupling between μ_1 and μ_2, and that transport happens only along transport rays (thanks to S. Szarek for this remark). Alternatively, one can use a perturbative argument based on the maximality of the integral $\int u f d\lambda$.

We refer to [50] for the details. □

As an application of this theorem, let us prove the reverse Cheeger inequality of Buser [26] and Ledoux [58], and in fact a refinement due to E. Milman [67]. In Section 2.3 above we saw another proof, using semi-group methods, of the following:

Proposition 31. *Let μ be a log-concave probability measure on $\mathbb{R}^n$ and $R > 0$. Assume that for any 1-Lipschitz function $u : \mathbb{R}^n \to \mathbb{R}$ there exists $\alpha \in \mathbb{R}$ with*

$$\int_{\mathbb{R}^n} |u(x) - \alpha| d\mu(x) \leq R. \tag{20}$$

(this is a weaker condition than requiring $C_P(\mu) \leq R^2$). Then for any measurable set $S \subseteq \mathbb{R}^n$ and $0 < \varepsilon < R$,

$$\mu(S_\varepsilon \setminus S) \geq c \cdot \frac{\varepsilon}{R} \cdot \mu(S) \cdot (1 - \mu(S)), \tag{21}$$

where $c > 0$ is a universal constant, and where S_ε is the ε-neighborhood of S. In particular the Cheeger constant of μ (see section 2.3) satisfies

$$\psi_\mu \lesssim R.$$

Proof. Denote $t = \mu(S) \in [0, 1]$ and set $f(x) = 1_S(x) - t$ for $x \in \mathbb{R}^n$. Then $\int f \, d\mu = 0$. We then consider the Monge transportation problem between $f_+ \, d\mu$ and $f_- \, d\mu$. Let u be a 1-Lipschitz function maximizing

$$\int_{\mathbb{R}^n} u f d\mu.$$

After adding a constant to u, we may assume that

$$\int_{\mathbb{R}^n} |u| d\mu \leq R.$$

By Theorem 29, we obtain a needle decomposition: measures $\{\mu_{\mathcal{I}}\}_{\mathcal{I}\in\Omega}$ on $\mathbb{R}^n$, and a measure ν on the space Ω of transport rays which yield a disintegration of measure. Observe that the equality

$$\int_{\Omega} \mu_{\mathcal{I}}(\mathbb{R}^n)\, \nu(d\mathcal{I}) = \mu(\mathbb{R}^n) = 1$$

implies in particular that for ν-almost every $\mathcal{I}$ the measure $\mu_{\mathcal{I}}$ is finite. We may normalize and assume that they are all probability measures. More precisely we can replace each of the measures $\mu_{\mathcal{I}}$ by $\mu_{\mathcal{I}}/\mu_{\mathcal{I}}(\mathbb{R}^n)$ and replace ν by the measure having density $\mathcal{I} \mapsto \mu_{\mathcal{I}}(\mathbb{R}^n)$ with respect to ν. Hence,

$$\int_{\Omega} \left(\int_{\mathcal{I}} |u|\, d\mu_{\mathcal{I}} \right) \nu(d\mathcal{I}) = \int_{\mathbb{R}^n} |u|\, d\mu \leq R.$$

Denote

$$B = \left\{ \mathcal{I} \in \Omega\, ;\, \int_{\mathcal{I}} |u|\, d\mu_{\mathcal{I}} \leq 2R \right\}.$$

By the Markov-Chebyshev inequality,

$$\nu(B) \geq 1/2. \tag{22}$$

For ν-almost all intervals $\mathcal{I} \in \Omega$ we know that $\int_{\mathcal{I}} f d\mu_{\mathcal{I}} = 0$, hence

$$\mu_{\mathcal{I}}(S) = t \cdot \mu_{\mathcal{I}}(\mathbb{R}^n) = t.$$

We would like to prove that for any $\mathcal{I} \in B$ and any $0 < \varepsilon < R$,

$$\mu_{\mathcal{I}}(S_\varepsilon \setminus S) \geq c \cdot \frac{\varepsilon}{R} \cdot t(1-t), \tag{23}$$

for a universal constant $c > 0$. Once (23) is proven, the bound (21) follows by integrating (23) with respect to ν and using (22), since

$$\mu(S_\varepsilon \setminus S) \geq \int_B \mu_{\mathcal{I}}(S_\varepsilon \setminus S)\, \nu(d\mathcal{I}) \geq \nu(B) \cdot c \cdot \frac{\varepsilon}{R} \cdot t(1-t) \geq \frac{c}{2} \cdot \frac{\varepsilon}{R} \cdot t(1-t).$$

What remains to be proven is a one-dimensional statement about log-concave measures: If $\eta = \mu_{\mathcal{I}}$ is a log-concave probability measure on $\mathbb{R}$ with $\int_{\mathbb{R}} |t| d\eta(t) \leq 2R$, then (23) holds true. This follows from Corollary 7 and a scaling argument. □

The same proof applies for any complete Riemannian manifold with non-negative Riemannian curvature. In fact, completeness in unneeded, the weaker geodesic-convexity assumption suffices here. There are quite a few other applications for this theorem, which helps reduce the task of proving an n-dimensional inequality to the task of proving a 1-dimensional inequality ("localization"). In a simply-connected space of constant sectional curvature, most of these applications – like reverse Hölder inequalities for polynomials – may also be proven using a localization method based on hyperplane bisections that go back to Payne and Weinberger [73], Gromov and Milman [42] and Kannan, Lovász and Simonovits [45]. Proposition 31 seems to be an exception, our proof requires the 1-Lipschitz guiding function.

Exercise 5 (reverse Hölder inequalities for polynomials)**.** Let X be a log-concave random vector in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most d. Then for any $0 < p \leq q$,

$$\|f(X)\|_q \leq C_{q,d} \cdot \|f(X)\|_p,$$

for some constant $C_{q,d}$ depending only on q and d.
Hint: In one dimension, following Bobkov [15], we may assume that f is a monic polynomial in one real variable, hence

$$f(X) = \prod_{i=1}^{d} (X - z_i)$$

for some $z_1, \ldots, z_d \in \mathbb{C}$. Consequently, by Hölder inequality and by Corollary 5,

$$\|f(X)\|_q = \left\| \prod_{i=1}^{d} (X - z_i) \right\|_q \leq \prod_{i=1}^{d} \|X - z_i\|_{dq} \leq \prod_{i=1}^{d} Cd(q+1) \|X - z_i\|_0 = (Cd(q+1))^d \|f(X)\|_0.$$

Now use needle decomposition to extend this to higher dimensions.

3.2 Isoperimetry and the Poincaré inequality

Recall that the Cheeger inequality [31] states that for any absolutely continuous probability measure on $\mathbb{R}^n$ satisfying some mild regularity assumptions,

$$C_P(\mu) \leq 4\psi_\mu^2. \tag{24}$$

The proof is sketched in section 2.3. Combining this with Proposition 31 we thus recover the aforementioned result by Buser, that $C_P(\mu)$ and ψ_μ^2 are within a constant factor of each other when μ is log-concave. Proposition 31 moreover implies that in the log-concave case, there exists a 1-Lipschitz function f such that

$$\psi_\mu^2 \leq C \cdot \mathrm{Var}_\mu(f).$$

This provides another proof of E. Milman's theorem (Corollary 21).

4 Bochner identities and curvature

In this lecture we discuss a technique that originated in Riemannian Geometry and connects the Poincaré inequality and Curvature. It started with the works of Bochner in the 1940s and also Lichnerowicz in the 1950s. The approach fits well with convex bodies and log-concave measures in high dimension. In a nutshell, the idea is to make local computations involving something like curvature, as well as integrations by parts, and then dualize and obtain Poincaré-type inequalities. This may sound pretty vague, let us explain what we mean.

Suppose that μ is an absolutely continuous log-concave probability measure in $\mathbb{R}^n$. Then μ is supported in an open, convex set $K \subseteq \mathbb{R}^n$ and it has a positive, log-concave density $\rho = e^{-\psi}$ in K. We will measure distances using the Euclidean distances in $\mathbb{R}^n$, but we will measure volumes using the measure μ. We thus look at the *weighted Riemannian manifold* or the *metric-measure space*

$$(K, |\cdot|, \mu).$$

Thus the Dirichlet energy of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ is

$$\|f\|^2_{\dot{H}^1(\mu)} = \int_K |\nabla f|^2 d\mu.$$

Indeed, we measure the length of the gradient with respect to the Euclidean metric, while we integrate with respect to the measure μ. As was already defined in Section 2.2, the Laplace-type operator associated with this measure-metric space is defined, initially for $u \in C_c^\infty(K)$, via

$$Lu = L_\mu u = \Delta u - \nabla\psi \cdot \nabla u = e^{\psi} div(e^{-\psi}\nabla u).$$

This reason for this definition is that for any smooth functions $u, v : \mathbb{R}^n \to \mathbb{R}$, with one of them compactly-supported in K,

$$\int_{\mathbb{R}^n} (Lu) v d\mu = -\int_{\mathbb{R}^n} [\nabla u \cdot \nabla v] e^{-\psi}\, dx.$$

and in particular

$$\langle -Lu, u\rangle_{L^2(\mu)} = \int_{\mathbb{R}^n} |\nabla u|^2 d\mu.$$

Thus L is a symmetric operator in $L^2(\mu)$, defined initially for $u \in C_c^\infty(K)$. It can have more than one self-adjoint extension, for example corresponding to the Dirichlet or Neumann boundary conditions when K is bounded. When discussing the Bochner technique, it is customary and possible to find ways to circumvent spectral theory of the operator L. Still, spectral theory helps us understand and form intuition, and we will at least quote the relevant spectral theory.

It will be convenient to make an (inessential) regularity assumption on μ, so as to avoid all boundary terms in all integrations by parts. We say that μ is a regular, log-concave measure in $\mathbb{R}^n$ if its density, denoted by $e^{-\psi}$, is smooth and positive in $\mathbb{R}^n$ and the following two requirements hold:

(i) Log-concavity amounts to ψ being convex, so $\nabla^2\psi \geq 0$ everywhere in $\mathbb{R}^n$. We require a bit more, that there exists $\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$,

$$\varepsilon \cdot \mathrm{Id} \leq \nabla^2\psi(x) \leq \frac{1}{\varepsilon} \cdot \mathrm{Id}. \tag{25}$$

(ii) The function ψ, as well as each of its partial derivatives, grows at most polynomially at infinity.

Exercise 6 (regularization process)**.** Begin with an arbitrary log-concave measure μ on $\mathbb{R}^n$, convolve it by a tiny Gaussian, and then multiply its density by $\exp(-\varepsilon|x|^2)$ for small $\varepsilon > 0$. Show that the resulting measure is regular, log-concave, with approximately the same covariance matrix, and that the Poincaré constant cannot jump down by much under this regularization process.

From now on, we assume that our probability measure μ is a regular, log-concave measure. It turns out that in this case, the operator L, initially defined on $C_c^\infty(\mathbb{R}^n)$, is essentially self-adjoint, positive semi-definite operator in $L^2(\mu)$ with a discrete spectrum. Its eigenfunctions $1 \equiv \varphi_0, \varphi_1, \ldots$ constitute an orthonormal basis, and the eigenvalues of $-L$ are

$$0 = \lambda_0(L) < \lambda_1(L) = \frac{1}{C_P(\mu)} \leq \lambda_2(L) \leq \ldots$$

with the eigenfunction corresponding to the trivial eigenvalue 0 being the constant function. The eigenfunctions are smooth functions in $\mathbb{R}^n$ that do not grow too fast at infinity: each function

$$\varphi_j e^{-\psi/2}$$

decays exponentially at infinity. Also $(\partial^\alpha \varphi_j)e^{-\psi/2}$ decays exponentially at infinity for any partial derivative α. This follows from known results on exponential decay of eigenfunctions of Schrödinger operators. The eigenvalues are given by the following infimum of Rayleigh quotients

$$\lambda_k(L) = \inf_{f \perp \varphi_0, \ldots, \varphi_{k-1}} \frac{\int_{\mathbb{R}^n} |\nabla f|^2 d\mu}{\int_{\mathbb{R}^n} f^2 d\mu}$$

where the infimum runs over all (say) locally-Lipschitz functions $f \in L^2(\mu)$. Since $\varphi_0 \equiv 1$, we indeed see that the first eigenfunction φ_1 saturates the Poincaré inequality for μ. For proofs of these spectral theoretic facts, see references in [54].

Let us return to Geometry. In Riemannian geometry, the Ricci curvature appears when we commute the Laplacian and the gradient. Analogously, here we have the easily-verified commutation relation

$$\nabla(Lu) = L(\nabla u) - (\nabla^2\psi)(\nabla u),$$

where $L(\nabla u) = (L(\partial^1 u), \ldots, L(\partial^n u))$. Hence the matrix $\nabla^2\psi$ corresponds to a curvature term, analogous to the Ricci curvature.

Proposition 32 (Integral Bochner's formula)**.** *For any* $u \in C_c^\infty(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} (Lu)^2 \ d\mu = \int_{\mathbb{R}^n} \left(\nabla^2 \psi\right) \nabla u \cdot \nabla u \ d\mu + \int_{\mathbb{R}^n} \|\nabla^2 u\|_{HS}^2 \, d\mu,$$

where $\|\nabla^2 u\|_{HS}^2 = \sum_{i=1}^n |\nabla \partial_i u|^2$.

Proof. Integration by parts gives

$$\begin{aligned}\int_{\mathbb{R}^n} (Lu)^2 \ d\mu &= -\int_{\mathbb{R}^n} \nabla(Lu) \cdot \nabla u \ d\mu \\ &= -\int_{\mathbb{R}^n} L(\nabla u) \cdot \nabla u \ d\mu + \int_{\mathbb{R}^n} \left[(\nabla^2 \psi) \nabla u \cdot \nabla u\right] \ d\mu \\ &= \sum_{i=1}^n \int_{\mathbb{R}^n} |\nabla \partial_i u|^2 \ d\mu + \int_{\mathbb{R}^n} \left(\nabla^2 \psi\right) \nabla u \cdot \nabla u \ d\mu.\end{aligned}$$

□

The assumption that u is compactly-supported was used in order to discard the boundary terms when integrating by parts. In fact, it suffices to know that u is μ-tempered. We say that u is μ-tempered if it is a smooth function, and $(\partial^\alpha u) e^{-\psi/2}$ decays exponentially at infinity for any partial derivative $\partial^\alpha u$. Any eigenfunction of L is μ-tempered. If f is μ-tempered, then so is Lf.

The following inequality from [54] is analogous to some investigations of Lichnerowicz [64]. It is concerned with distributions that are more log-concave than a Gaussian distribution, in the sense that their logarithmic Hessian is uniformly bounded by that of the Gaussian.

Theorem 33 (improved log-concave Lichnerowicz inequality)**.** *Let* $t > 0$ *and assume that* $\nabla^2 \psi(x) \geq t$ *for all* $x \in \mathbb{R}^n$*. Then,*

$$C_P(\mu) \leq \sqrt{\|\mathrm{Cov}(\mu)\|_{op} \cdot \frac{1}{t}}.$$

Equality in Theorem 33 is attained when μ is a Gaussian measure, with any covariance matrix. Indeed in that case $C_P(\mu)$ and $\|\mathrm{Cov}(\mu)\|_{op}$ coincide, and they also coincide with the inverse lower bound on the Hessian of the potential. Write γ_s for the law of distribution of a Gaussian random vector of mean zero and covariance matrix $s \cdot \mathrm{Id}$ in $\mathbb{R}^n$. Then γ_s satisfies the assumptions of Theorem 33 for $t = 1/s$ while $C_P(\gamma_s) = \|\mathrm{Cov}(\gamma_s)\|_{op} = s$.

Proof of Theorem 33. Denote $f = \varphi_1$, the first eigenfunction, normalized so that $\|f\|_{L^2(\mu)} = 1$. Set $\lambda = 1/C_P(\mu)$. By the Bochner formula and the Poincaré inequality for $\partial^i f$ $(i = 1, \ldots, n)$,

$$\begin{aligned}\lambda^2 = \int_{\mathbb{R}^n} (Lf)^2 d\mu &= \int_{\mathbb{R}^n} [(\nabla^2 \psi) \nabla f \cdot \nabla f] d\mu + \int_{\mathbb{R}^n} \|\nabla^2 f\|_{HS}^2 d\mu \\ &\geq t \int_{\mathbb{R}^n} |\nabla f|^2 d\mu + \lambda \left[\int_{\mathbb{R}^n} |\nabla f|^2 d\mu - \left|\int_{\mathbb{R}^n} \nabla f d\mu\right|^2\right] \\ &= (t + \lambda) \cdot \lambda - \lambda \left|\int_{\mathbb{R}^n} \nabla f d\mu\right|^2. \end{aligned} \tag{26}$$

Therefore the first eigenfunction has a "preferred direction", i.e.,

$$\left| \int_{\mathbb{R}^n} \nabla f d\mu \right|^2 \geq t. \tag{27}$$

We remark that in the general case, under log-concavity assumptions it is known that $\int_{\mathbb{R}^n} \nabla f d\mu \neq 0$, see [49], and this leads to a bound on the dimension of the first eigenspace. The lower bound (27) is a quantitative version, relying on the assumption of a uniform lower bound on the log-concavity. Using that the i^{th} coordinate of ∇f is $\nabla f \cdot \nabla x_i$ and integrating by parts we have

$$\int_{\mathbb{R}^n} \nabla f d\mu = -\int_{\mathbb{R}^n} (Lf) x d\mu = \lambda \int_{\mathbb{R}^n} f x d\mu$$

Since $\int f d\mu = 0$, by Cauchy-Schwartz, for some $\theta \in S^{n-1}$,

$$\begin{aligned}\left| \int_{\mathbb{R}^n} \nabla f d\mu \right| &= \int_{\mathbb{R}^n} \langle \nabla f, \theta \rangle d\mu = \lambda \int_{\mathbb{R}^n} f(x) \langle x, \theta \rangle \, \mu(dx) \\ &\leq \lambda \|f\|_{L^2(\mu)} \cdot \sqrt{\mathrm{Cov}(\mu)\theta \cdot \theta} \leq \lambda \|\mathrm{Cov}(\mu)\|_{op}.\end{aligned}$$

This expression is at least t, and the theorem follows. □

Since $\|\mathrm{Cov}(\mu)\|_{op} \leq C_P(\mu)$, we deduce from Theorem 33 that

$$C_P(\mu) \leq \frac{1}{t}. \tag{28}$$

Inequality (28) is sometimes referred to as the log-concave Lichnerowicz inequality. Therefore the bound in Theorem 33 is a geometric average of the Lichnerowicz bound and the conjectural KLS bound.

The Bochner identity has quite a few additional applications in the study of log-concave measures, beyond the improved log-concave Lichnerowicz inequality. Especially if one introduces the semigroup $(e^{tL})_{t \geq 0}$ associated with the operator L (see e.g. Ledoux [59]), as we saw in Section 2.2. Yet even simple integrations by parts and duality arguments based on the Bochner identity lead to non-trivial conclusions. One example is the Brascamp-Lieb inequality [24] from the 1970s:

Theorem 34 (Brascamp-Lieb)**.** *For any C^1-smooth $f \in L^2(\mu)$,*

$$\mathrm{Var}_\mu(f) \leq \int_{\mathbb{R}^n} \left(\nabla^2 \psi \right)^{-1} \nabla f \cdot \nabla f \, \mu(dx),$$

where $\mathrm{Var}_\mu(f) = \int_{\mathbb{R}^n} (f - E)^2 \, \mu(dx)$, *and* $E = \int_{\mathbb{R}^n} f d\mu$.

Proof. We will only prove this inequality for regular, log-concave measures, though it holds true under weaker regularity assumptions. The space of all μ-tempered functions is denoted by $\mathcal{F}_\mu$. It is clearly a dense subspace of $L^2(\mu)$ and in fact its image under L is dense in

$$\varphi_0^\perp = \left\{ g \in L^2(\mu) \, ; \, \int_{\mathbb{R}^n} g d\mu = 0 \right\}.$$

Indeed, the image contains all finite linear combinations of all eigenfunctions $\varphi_1, \varphi_2, \dots$ (without φ_0) which is dense in H. Assume $\int f \ d\mu = 0, \varepsilon > 0$ and pick $u \in \mathcal{F}_\mu$ such that

$$\|Lu - f\|_{L^2(\mu)} < \varepsilon.$$

Then,

$$\begin{aligned}\mathrm{Var}_\mu(f) = \|f\|^2_{L^2(\mu)} &= \|Lu - f\|^2_{L^2(\mu)} + 2\int fLu \ d\mu - \int (Lu)^2 \ d\mu \\ &\le \varepsilon^2 - 2\int \nabla f \cdot \nabla u \ d\mu - \int (\nabla^2\psi)\nabla u \cdot \nabla u \ d\mu \\ &\le \varepsilon^2 + \int (\nabla^2\psi)^{-1}\nabla f \cdot \nabla f \ d\mu,\end{aligned}$$

where we have used the fact that

$$\int (Lu)^2 \ d\mu \ge \int (\nabla^2\psi)\nabla u \cdot \nabla u \ d\mu,$$

which follows from Bochner's formula and

$$-2x \cdot y - Ax \cdot x \le A^{-1}y \cdot y \Longleftrightarrow |\sqrt{A}x + \sqrt{A^{-1}}y|^2 \ge 0.$$

The desired inequality follows by letting ε tend to zero. □

Remark. The Brascamp-Lieb inequality is an infinitesimal version of the Prékopa-Leindler inequality. Suppose that $f_0, f_1 : \mathbb{R}^n \to [0, \infty)$ are integrable, log-concave functions and

$$f_t(x) = \sup_{x=(1-t)y+yz} f_0(y)^{1-t} f_1(z)^t.$$

The Prékopa-Leindler inequality implies that $\log \int_{\mathbb{R}^n} f_t$ is concave in t. The second derivative in t is non-negative, and this actually amounts to the Brascamp-Lieb inequality. Thus the Brascamp-Lieb inequality is yet another incarnation of the Brunn-Minkowski inequality.

We say that a function ψ on the orthant $\mathbb{R}^n_+$ is p-convex if $\psi(x_1^{1/p}, \dots, x_n^{1/p})$ is a convex function of $(x_1, \dots, x_n) \in \mathbb{R}^n_+$.

Corollary 35. *Let μ be a probability measure in the orthant $\mathbb{R}^n_+$, set $e^{-\psi} = d\mu/dx$ and assume that ψ is p-convex for $p = 1/2$. Then for any C^1-smooth function $f \in L^2(\mu)$,*

$$\mathrm{Var}_\mu(f) \le 4 \int_{\mathbb{R}^n} \sum_{i=1}^n x_i^2 |\partial_i f|^2 \ \mu(dx).$$

For general $p > 1$, replace the coefficient 4 by $p^2/(p-1)$.

Proof. Change variables and use the Brascamp-Lieb inequality. Denote $\frac{d\mu}{dx} = e^{-\psi}$. Then for

$$\pi(x_1, \cdots, x_n) = (x_1^2, \cdots, x_n^2),$$

the function $\psi(\pi(x))$ is convex. Set

$$\varphi(x) = \psi(\pi(x)) - \sum_{i=1}^{n} \log(2x_i).$$

Then π^{-1} pushes-forward μ to the measure with density $e^{-\varphi}$. Moreover,

$$\nabla^2 \varphi(x) \geq \nabla^2 \left(-\sum_{i=1}^{n} \log(2x_i) \right) = \begin{pmatrix} \frac{1}{x_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_n^2} \end{pmatrix} > 0,$$

and therefore

$$\left(\nabla^2 \varphi(x)\right)^{-1} \leq \begin{pmatrix} x_1^2 & 0 & \cdots & 0 \\ 0 & x_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n^2 \end{pmatrix}.$$

Set $g(x) = f(\pi(x))$. By the Brascamp-Lieb inequality,

$$\mathrm{Var}_{e^{-\varphi}}(g) \leq \int_{\mathbb{R}_+^n} \left[\left(\nabla^2 \varphi\right)^{-1} \nabla g \cdot \nabla g \right] e^{-\varphi(x)} \, dx \leq \int_{\mathbb{R}_+^n} \sum_{i=1}^{n} x_i^2 |\partial_i g(x)|^2 e^{-\varphi(x)} \, dx.$$

The corollary follows since

$$\mathrm{Var}_{e^{-\varphi}}(g) = \mathrm{Var}_{e^{-\psi}}(f).$$

and since when $y = \pi(x) = (x_1^2, \cdots, x_n^2)$ we have

$$x_i \partial_i g(x) = 2 y_i \partial_i f(y).$$

□

Exercise 7. If $\psi : \mathbb{R}_+^n \to \mathbb{R}$ is convex and increasing in all of the coordinate directions, then ψ is p-convex for $p = 1/2$, i.e., $\psi(x_1^2, \dots, x_n^2)$ is convex in the orthant.

A function $\psi : \mathbb{R}^n \to \mathbb{R}$ is invariant under coordinate reflections (a.k.a. unconditional) if

$$\psi(x_1, \dots, x_n) = \psi(|x_1|, \dots, |x_n|) \qquad \text{for all } x \in \mathbb{R}^n.$$

If ψ is moreover convex, then $\psi|_{\mathbb{R}_+^n}$ is increasing in all coordinate directions. Similarly a random vector is called unconditional if its law is invariant under reflection by coordinate hyperplanes. When the vector has a density this amounts to saying the density is unconditional in the above sense. The following thin-shell bound is from [49].

Corollary 36. *Suppose that X is a random vector that is log-concave, isotropic and unconditional in $\mathbb{R}^n$. Then,*

$$\mathrm{Var}(|X|^2) \leq Cn.$$

Proof. According to the exercise the density of X is of the form $\mathrm{e}^{-\psi}$, where ψ is p-convex for $p = 1/2$. Corollary 35 applies and we get

$$\mathrm{Var}(|X|^2) \leq 4\sum_{i=1}^{n} \mathbb{E}X_i^2(2X_i)^2 = 16\sum_{i=1}^{n} \mathbb{E}X_i^4 \lesssim \sum_{i=1}^{n} (\mathbb{E}X_i^2)^2 = n.$$

where we used reverse Hölder inequalities in the last passage. □

It should be noted that the result is optimal, in the sense that there exist unconditional isotropic log-concave random vectors X for which $\mathrm{Var}(|X|^2)$ is of order n, see the exercise below. We also remark that as of October 2024, the state of affairs is that the KLS conjecture is still open already in the particular case of unconditional convex bodies. A logarithmic bound for the Poincaré constant in this case is known for years, see [49], and it is subsumed by recent bounds for the general case.

Exercise 8. If X is a standard Gaussian vector in $\mathbb{R}^n$,

$$\mathrm{Var}(|X|^2) = 2n.$$

5 Gaussian localization

In the previous section we discussed localization of a log-concave measure into *needles*, one-dimensional segments. We proceed by discussing Gaussian localization, decomposing the given measure into a mixture of measures, each of which involves multiplying the given measure by a Gaussian. The Gaussians bring with them a wealth of connections and elegant formulae, as we see below. The method was invented by Ronen Eldan [35] and it is coined *Eldan's Stochastic Localization*. We first present a rather degenerate case of Eldan's method, in which the time parameter is somewhat fixed, so that the method does not require stochastic processes.

Let Z be a standard Gaussian random vector in $\mathbb{R}^n$, of mean zero and identity covariance matrix Id. Recall that for $s > 0$ we write γ_s for the density of $\sqrt{s} \cdot Z$. Let X be a log-concave random vector in $\mathbb{R}^n$ independent of Z, with density ρ. For $s \geq 0$ consider the random vector

$$Y_s = X + \sqrt{s}Z$$

whose density is $\rho * \gamma_s$.

One could think of (Y_s) as a process parameterized by s, perhaps as a Brownian motion starting at the initial distribution of X. This point of view, with the time reversal $t = 1/s$, is emphasized in Section 6. In the present lecture do not consider a stochastic process parameterized by s, and view $s > 0$ as a parameter whose value will be fixed later on. One of the simplest examples of Gaussian localization of the probability density ρ is given by the following:

Proposition 37. *Fix $s > 0$. For each $y \in \mathbb{R}^n$, consider the probability density*

$$\rho_{s,y}(x) = \frac{\rho(x)\gamma_s(x-y)}{\rho * \gamma_s(y)},$$

which we view as a localized "Gaussian needle" or "Gaussian piece" relative to ρ. Then the original density ρ is a certain average of these Gaussian needles:

$$\rho = \mathbb{E}\rho_{s,Y_s}.$$

One says that this is a disintegration of ρ into the localized Gaussian pieces $(\rho_{s,y})_{y\in\mathbb{R}^n}$.

Proof. The joint density of (X, Y_s) in $\mathbb{R}^n \times \mathbb{R}^n$ is

$$(x, y) \mapsto \rho(x)\gamma_s(y - x).$$

The family of densities $\rho_{s,y}$ give us the conditional distribution of X with respect to Y_s. That is, for any test function $f(x, y)$,

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x,y)\rho(x)\gamma_s(y-x)dxdy = \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} f(x,y)\rho_{s,y}(x)dx\right]\rho * \gamma_s(y)dy$$

In particular, if the function $f(x, y)$ depends only on x, we get

$$\int_{\mathbb{R}^n} f\rho = \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} f\rho_{s,y}\right]\rho * \gamma_s(y)dy = \mathbb{E}\int_{\mathbb{R}^n} f\rho_{s,Y_s}. \qquad \square$$

From the proof of Proposition 37 we see that the densities $\rho_{s,y}$ give us the conditional distribution of X with respect to Y_s. The conditional expectation operator is denoted by

$$Q_s f(y) = \int_{\mathbb{R}^n} f \rho_{s,y},$$

whenever the integral converges. Thus

$$Q_s f(Y_s) = \mathbb{E}\left[f(X) | Y_s\right].$$

Assume that the original density ρ is log-concave. Then each of the elements $\rho_{s,y}$ in the decomposition is more log-concave than the Gaussian γ_s. We have thus expressed our log-concave density as a mixture of measures that are uniformly log-concave. This decomposition is determined by the choice of the parameter $s > 0$.

The critical value of s turns out to be $s \sim C_P(X)$. Roughly speaking, for much smaller values of s, we decompose into highly localized measures, maybe even resembling Dirac masses. For much larger values of s the decomposition is trivial for another reason: the localized pieces resemble the original measure. Abbreviate

$$\rho_s = \rho_{s,Y_s},$$

a random probability density. Recall that $\mathbb{E}\rho_s = \rho$ by Proposition 37. As usual, for a function f on $\mathbb{R}^n$ we write

$$\mathrm{Var}_{\rho_s}(f) = \int_{\mathbb{R}^n} f^2 \rho_s - \left(\int_{\mathbb{R}^n} f \rho_s \right)^2,$$

provided that the integrals converge. Similarly, we also write $\mathrm{Var}_\rho(f) = \mathrm{Var} f(X)$. Then by the law of total variance,

$$\mathrm{Var} f(X) = \mathbb{E}\mathrm{Var}(f(X)|Y_s) + \mathrm{Var}(\mathbb{E}(f(X)|Y_s)) = \mathbb{E}\mathrm{Var}_{\rho_s}(f) + \mathrm{Var}(Q_s f(Y_s)). \tag{29}$$

When $s \gtrsim C_P(X)$, it is the first summand that is dominant:

Lemma 38. *Assume that X is log-concave. For any $s > 0$ and a function f on $\mathbb{R}^n$ with $\mathbb{E} f^2(X) < \infty$,*

$$\mathbb{E}\mathrm{Var}_{\rho_s}(f) \le \mathrm{Var}_\rho(f) \le \left(2 + \frac{C_P(X)}{s}\right) \mathbb{E}\mathrm{Var}_{\rho_s}(f).$$

Proof. We need to show that $\mathrm{Var} Q_s f(Y_s)$ is not much larger than $\mathbb{E}\mathrm{Var}_{\rho_s}(f)$. To this end, we will use the Poincaré inequality for the random vector Y_s. By the subadditivity property of the Poincaré constant (see exercise 2

$$C_P(Y_s) = C_P(X + \sqrt{s} Z) \le C_P(X) + C_P(\sqrt{s} Z) = C_P(X) + s.$$

Hence

$$\mathrm{Var}(Q_s f(Y_s)) \le (C_P(X) + s) \cdot \mathbb{E}|\nabla Q_s f(Y_s)|^2.$$

Recall that

$$Q_s f(y) = \int_{\mathbb{R}^n} \rho_{s,y}(x) f(x) dx = \int_{\mathbb{R}^n} \frac{\rho(x)\gamma_s(x-y)}{\rho * \gamma_s(y)} f(x) dx.$$

Differentiating a Gaussian is easy, we have $\nabla \gamma_s(x) = -\gamma_s(x) \cdot x/s$. It follows that

$$\nabla Q_s f(y) = \int_{\mathbb{R}^n} \frac{x - a_s}{s} \rho_{s,y}(x) f(x) dx,$$

where $a_s = a_{s,y} = \int_{\mathbb{R}^n} x \rho_{s,y}(x) dx$ is the barycenter of the local measure $\rho_{s,y}$. Write $A_s = A_{s,y} = \mathrm{Cov}(\rho_{s,y})$. By the Cauchy-Schwartz inequality, for $\theta \in S^{n-1}$,

$$\begin{aligned}
\nabla Q_s f(y) \cdot \theta &= \int_{\mathbb{R}^n} \frac{(x - a_s) \cdot \theta}{s} \rho_{s,y}(x) f(x) dx \\
&\leq \frac{1}{s} \sqrt{\int_{\mathbb{R}^n} |(x - a_s) \cdot \theta|^2 \rho_{s,y}(x) dx} \sqrt{\mathrm{Var}_{\rho_{s,y}}(f)} \\
&\leq \frac{1}{s} \sqrt{\|A_s\|_{op}} \cdot \sqrt{\mathrm{Var}_{\rho_{s,y}}(f)}.
\end{aligned}$$

Then by taking the supremum over $\theta \in S^{n-1}$,

$$\mathrm{Var}(Q_s f(Y_s)) \leq \frac{C_P(X) + s}{s^2} \cdot \mathbb{E}\left(\|A_s\|_{op} \mathrm{Var}_{\rho_s}(f) \right).$$

However, the random probability density ρ_s is always more log-concave than the Gaussian γ_s, and hence $A_s \leq s \cdot \mathrm{Id}$. Consequently,

$$\mathrm{Var}(Q_s f(Y_s)) \leq \frac{C_P(X) + s}{s} \cdot \mathbb{E}\mathrm{Var}_{\rho_s}(f).$$

This, together with (29), proves the proposition. □

To summarize, for $s \gtrsim C_P(\mu)$, the local measure ρ_s is typically close enough to the original measure, so the variance of any fixed function with respect to ρ is roughly the averaged variance with respect to ρ_s.

Remark. By differentiating with respect to s, one may improve upon Proposition 38 in two respects. First, it turns out that log-concavity is actually not needed in Proposition 38. It is proven in Klartag and Ordentlich [56] that for any random vector X and a function f with $\mathbb{E} f^2(X) < \infty$,

$$\mathrm{Var}_\rho(f) \leq \left(1 + \frac{C_P(X)}{s}\right) \mathbb{E}\mathrm{Var}_{\rho_s}(f). \tag{30}$$

This is a better bound than that of Lemma 38.

Corollary 39. *For any $s > 0$, setting $\alpha = s/C_P(X)$,*

$$C_P(X) \leq C\left(1 + \frac{1}{\alpha}\right) \cdot \mathbb{E} C_P(\rho_s),$$

where $C > 0$ is a universal constant.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a 1-Lipschitz function with

$$\mathrm{Var}_\mu(f) \geq c \cdot C_P(X),$$

whose existence in guaranteed by Corollary 21 due to E. Milman. By Proposition 38 and the Poincaré inequality,

$$\begin{aligned}
\mathrm{Var}_\mu(f) &\leq \left(2 + \frac{1}{\alpha}\right) \mathbb{E}\mathrm{Var}_{\rho_s}(f) \\
&\leq \left(2 + \frac{1}{\alpha}\right) \mathbb{E}\left(C_P(\rho_s) \cdot \int_{\mathbb{R}^n} |\nabla f|^2 \rho_s\right) \\
&\leq \left(2 + \frac{1}{\alpha}\right) \mathbb{E} C_P(\rho_s). \qquad \square
\end{aligned}$$

Thus, in order to bound the Poincaré constant of X, we may apply Gaussian localization with $s \gtrsim C_P(\mu)$ and try to bound the Poincaré constant of ρ_s. An advantage of ρ_s over ρ is that ρ_s is more log-concave than the Gaussian γ_s. Hence, by the improved log-concave Lichnerowicz inequality, which is Theorem 33 above,

$$C_P(\rho_s) \leq \sqrt{s \cdot \|A_s\|_{op}}$$

where we recall that $A_s = \mathrm{Cov}(\rho_s)$. Therefore, Corollary 39 leads to another corollary:

Corollary 40. *For any $s > 0$,*

$$C_P(X) \leq C\left(1 + \frac{C_P(X)}{s}\right) \cdot \sqrt{\mathbb{E}\|A_s\|_{op} \cdot s}.$$

What do we know about $\mathbb{E}\|A_s\|_{op}$? Assume from now on that X is log-concave and isotropic, so for large $s > 0$ we might expect A_s to be roughly $\mathrm{Cov}(X) = \mathrm{Id}$. However, the operator norm involves a supremum, and this complicates matters. The evolution of the operator norm of the covariance matrix is analyzed in great detail in Section 7 using stochastic processes and computations involving 3-tensors, leading to the following estimate.

Theorem 41. *Define*

$$s_0 = \min\{s > 0\,;\, \forall r > s,\ \mathbb{E}\|A_r\|_{op} \leq 5\}.$$

Then,

$$s_0 \leq C \log^2(n+1) \tag{31}$$

where $C > 0$ is a universal constant. This bound utilizes the improved Lichnerowicz inequality, proven only recently. A slightly older bound that suffices here (e.g. [53, 55]) is

$$s_0 \leq C \log(n+1) \cdot \sup C_P(\mu)$$

where the supremum runs over all isotropic, log-concave probability measures μ on $\mathbb{R}^n$.

Moreover, $s_0 \geq c \log n$ in some examples, in particular when $1 + X_1, \dots, 1 + X_n$ are i.i.d. Exponential random variables with parameter 1.

One could conjecture that stochastic processes and pathwise analysis of are not essential for the proof of Theorem 41, and that an analytic proof is possible to find. There are other applications of stochastic localization which seem to rely heavily on pathwise analysis (e.g., the complex waist inequalities in [51]). By using Theorem 41 and Corollary 40 with $s = C\log^2(n+1)$ we thus arrive at

Corollary 42 ("best known bound for KLS")**.** *For any isotropic, log-concave random vector X in $\mathbb{R}^n$,*

$$C_P(X) \leq C\log(n+1) \tag{32}$$

where $C > 0$ is a universal constant.

Proof. We have

$$C_P(X) \leq C\left(1 + \frac{C_P(X)}{\log^2(n+1)}\right) \cdot \sqrt{\log^2(n+1)},$$

which implies (32). □

6 A dynamic perspective on Gaussian localization

Formally when s tends to ∞, the variable $X + \sqrt{s}G$ becomes independent of X, so the conditional law of X given $X + \sqrt{s}G$ tends to the law of X. In this section we will study the dynamic of this measure-valued process as time s evolves.

6.1 The Eldan equation

The process solves a certain stochastic differential equation which was first considered by Eldan and which we present now. We are given a probability measure on $\mathbb{R}^n$, and a standard Brownian motion (W_t) on $\mathbb{R}^n$. We consider the following infinite system of SDE whose unknown is the family (p_t) of functions from $\mathbb{R}^n$ to $\mathbb{R}_+$:

$$\begin{cases} p_0(x) = 1 \\ dp_t(x) = p_t(x)\,(x - a_t) \cdot dW_t, \end{cases}$$

where a_t is the barycenter $p_t(x)\mu(dx)$, namely

$$a_t = \frac{\int_{\mathbb{R}^n} x \cdot p_t(x)\,\mu(dx)}{\int_{\mathbb{R}^n} p_t(x)\,\mu(dx)}.$$

Note that we have only one Brownian motion (W_t) which is used for every x. Actually in Eldan's original paper [35] the equation is slightly more intricate than that. Here we consider the simplified version that was introduced by Lee and Vempala [63].

Since we have an equation for each x and they are all coupled together by the condition on the barycenter, it is not at all clear that such a process should actually exists. Let us leave that matter aside for now, we will come back to that later on. Let us also take for granted the fact that $p_t(x) > 0$ for all t, almost surely. The barycenter condition then ensures that the total mass of $p_t\,d\mu$ remains constant. Indeed, at least formally we have

$$d \int_{\mathbb{R}^n} p_t(x)\,\mu(dx) = \int_{\mathbb{R}^n} dp_t(x)\,\mu(dx) = \left(\int_{\mathbb{R}^n} (x - a_t) p_t(x)\,\mu(dx) \right) \cdot dW_t,$$

which is 0 by definition of a_t. Therefore $p_t\,d\mu$ is a random probability measure for all time, and we call that measure μ_t from now on. The second feature is that $p_t(x)$ is a martingale for all x. In particular $\mathbb{E}p_t(x) = p_0(x) = 1$ for all x. Therefore the random measure μ_t equals μ on average

$$\mathbb{E}\mu_t = \mu.$$

The third observation is that the equation

$$dp_t(x) = p_t(x)(x - a_t) \cdot dW_t$$

can be solved explicitly. Indeed applying Itô's formula to $\log p_t(x)$ we get

$$d \log p_t(x) = (x - a_t) \cdot dW_t - \frac{1}{2}|x - a_t|^2\,dt,$$

hence

$$p_t(x) = \exp\left(\int_0^t (x-a_s)\cdot dW_s - \frac{1}{2}\int_0^t |x-a_s|^2\, ds\right) = \exp\left(c_t + x\cdot\theta_t - \frac{t}{2}|x|^2\right),$$

where (c_t) and (θ_t) are certain random processes not depending on x. This shows that the density p_t of μ_t with respect to μ is just a certain Gaussian factor. The linear term and the normalizing constant are random but the quadratic term is deterministic, equal to $\frac{t}{2}|x|^2$. As a result if the original measure was log-concave then the measure μ_t is t-uniformly log-concave, almost surely. The process becomes more and more *peaked* as t grows. For this reason Eldan coined the name *stochastic localization process*. It allows us to write a log-concave measure as a mixture of t-uniformly log-concave measures. Moreover this mixture is constructed by solving a certain stochastic differential equation, so that its behavior over time can be somehow controlled using Itô's formula.

6.2 Proper construction of a solution

We will now give a rigorous construction of the stochastic localization process. As we said earlier this process was introduced by Eldan [35] (a variant of it actually), it was used in a number of subsequent works [63, 32, 55]. The construction that we give here is somewhat original, but very much inspired by Klartag-Putterman [57].

Start with a standard n-dimensional Brownian motion (θ_t) defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)$. This is an odd name for a Brownian motion, you'll see the reason for this choice shortly. Observe that for every fixed $x \in \mathbb{R}^n$ the process (E_t) given by

$$E_t = \exp\left(x\cdot\theta_t - \frac{t}{2}|x|^2\right)$$

is a martingale. Indeed, since the Brownian motion has independent and stationary increments, for every $s \leq t$, the ratio E_t/E_s is independent of whatever happens before times s and has expectation 1. Using Fubini, we deduce that given a test function f, the process (N_t) given by

$$N_t = \int_{\mathbb{R}^n} f(x)\cdot\exp\left(x\cdot\theta_t - \frac{t}{2}|x|^2\right)\,\mu(dx). \tag{33}$$

also is a martingale. In particular its expectation is what we have at time 0, namely $\int_{\mathbb{R}^n} f\, d\mu$. Let μ_t be the random probability measure on $\mathbb{R}^n$ given by

$$\mu_t(dx) = \frac{1}{D_t}\exp\left(x\cdot\theta_t - \frac{t}{2}|x|^2\right)\,\mu(dx), \tag{34}$$

where D_t is the normalization constant, namely

$$D_t = \int_{\mathbb{R}^n} \exp\left(x\cdot\theta_t - \frac{t}{2}|x|^2\right)\,\mu(dx) \tag{35}$$

We can then rewrite N_t as

$$N_t = D_t \cdot \int_{\mathbb{R}^n} f(x)\, d\mu_t. \tag{36}$$

We will interpret the normalizing factor D_t as a change in the probability space.

Fix a large but finite time horizon T. Since the process $(D_t)_{t\leq T}$ is a positive martingale with expectation 1, we can define a new probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by saying that $\mathbb{Q}$ has density D_T with respect to $\mathbb{P}$. Then it is easy to see that a process (X_t) defined on $[0,T]$ is a $\mathbb{Q}$-martingale if and only if the process $(X_t D_t)$ is a $\mathbb{P}$-martingale. Recall that the process (N_t) given by (33) was a $\mathbb{P}$-martingale. In view of (36) we thus the following.

Fact 43. *For any test function f, the process (M_t) given by $M_t = \int_{\mathbb{R}^n} f\, d\mu_t$ is a $\mathbb{Q}$-martingale.*

Getting an Itô equation for this process is a little more involved. It relies on the Girsanov change of measure formula which we spell out now.

Proposition 44 (Girsanov change of measure)**.** *If X is a $\mathbb{P}$-local martingale on $[0,T]$ then the process $\widetilde{X}$ given by*

$$\widetilde{X}_t = X_t - \int_0^t \frac{d\langle X, D\rangle_s}{D_s}$$

is a $\mathbb{Q}$-local martingale on $[0,T]$. Moreover, $\widetilde{X}$ and X have the same quadratic variation. In particular if X is a $\mathbb{P}$-Brownian motion on $[0,T]$ then $\widetilde{X}$ is a $\mathbb{Q}$-Brownian motion on $[0,T]$.

Remark 45. The bracket denotes the quadratic covariation of continuous semimartingales. Note that the quadratic variation under $\mathbb{P}$ is the same as the quadratic variation under $\mathbb{Q}$. Indeed, quadratic variation is defined as the limit in probability of sums of squared increments along partitions of the interval whose mesh sizes tend to 0. This is easily seen to be left unchanged by an absolutely continuous change of probability measure.

Remark 46. In the statement the process X is $\mathbb{R}$-valued but the result also works for vector valued martingales by applying it to each coordinate.

Proof. This is a very standard tool in stochastic calculus, we only give a very brief sketch of proof and refer to [76, section IV.38] for more details. This amounts to proving that $\widetilde{X}D$ is a $\mathbb{P}$-martingale. But, from Itô's integration by parts formula we get

$$\begin{aligned} d(\widetilde{X}D) &= (d\widetilde{X})D + \widetilde{X}(dD) + d\langle \widetilde{X}, D\rangle \\ &= (dX)D - d\langle X, D\rangle + \widetilde{X}(dD) + d\langle X, D\rangle. \end{aligned}$$

The quadratic covariation of X and D thus cancels out and we are left with martingale increments only. □

Coming back to our situation, we see that the change of measure is of the form

$$D_t = \exp(\phi(t, \theta_t)) \tag{37}$$

where $\phi\colon \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ is the function given by

$$\phi(t,\theta) = \log\left(\int_{\mathbb{R}^n} \exp\left(\langle x, \theta\rangle - \frac{t}{2}|x|^2 \right) \mu(dx) \right). \tag{38}$$

This is not quite essential but let us assume for simplicity that $\mathrm{e}^{x\cdot\theta}$ is μ-integrable for all $\theta \in \mathbb{R}^n$ in which case ϕ is smooth on $[0,\infty) \times \mathbb{R}^n$. From Itô's formula we get

$$dD_t = D_t \, \nabla\phi(t,\theta_t) \cdot d\theta_t.$$

Here and in the sequel, ∇ and Δ always mean gradient and Laplacian with respect to the space variable. The derivative with respect to the time variable will be denoted ∂_t. Then from Gisanov, we see that the process (W_t) given by

$$\begin{aligned} dW_t &= d\theta_t - \frac{d\langle \theta_t, D_t\rangle}{D_t} \\ &= d\theta_t - \nabla\phi(t,\theta_t)\, dt \end{aligned}$$

is a $\mathbb{Q}$-Brownian motion. We rewrite this equation as

$$d\theta_t = dW_t + \nabla\phi(t,\theta_t)\, dt. \tag{39}$$

We are now in a position to prove the following.

Fact 47. *The Itô derivative of the $\mathbb{Q}$-martingale $M_t = \int_{\mathbb{R}^n} f \, d\mu_t$ is given by*

$$dM_t = \left(\int_{\mathbb{R}^n} f(x)(x - a_t)\, d\mu_t \right) \cdot dW_t,$$

where $a_t = \int_{\mathbb{R}^n} x \, d\mu_t$ is the barycenter of μ_t.

Proof. First of all, by differentiating (38) under the integral sign, we obtain $\nabla\phi(t,\theta_t) = a_t$. We see $M_t = \int f \, d\mu_t$ as a function of t and θ_t, denoted $F(t,\theta_t)$. By Itô's formula and (39), we have

$$dM_t = \nabla F(t,\theta_t) \cdot (dW_t + \nabla\phi(t,\theta_t)\, dt) + \frac{1}{2}\Delta F(t,\theta_t)\, dt + \partial_t F(t,\theta_t)\, dt.$$

The gradient of F is

$$\nabla F(t,\theta_t) = \int_{\mathbb{R}^n} f(x)(x - a_t)\, d\mu_t.$$

Moreover, since we have seen above that $(F(t,\theta_t))$ is a martingale for some filtration for which (W_t) is also a martingale it must be the case that the dt part above cancels out. It can indeed be checked that F satisfies the following PDE

$$\partial_t F = -\nabla F \cdot \nabla\phi - \frac{1}{2}\Delta F.$$

This concludes the proof of the fact. □

Remark 48. Strictly speaking this only gives a construction of the process (μ_t) on a bounded time interval $[0, T]$. This will be sufficient for our needs but let us note that one could extend this construction to the whole half-line by some abstract argument à la Carathéodory. Beware though that the change of measure is only absolutely continuous when we restrict our processes to a bounded time interval.

As a byproduct of this construction we obtain a simple description of the law of the process (θ_t). This observation is not present in the works of Eldan, Lee-Vempala, and Chen. Its first explicit mention is in the paper of Klartag and Putterman.

Proposition 49. *The process (θ_t) has the same law as the process $(tX + W_t)$, where (W_t) is a standard Brownian motion, and X is a random vector having law μ independent of (W_t).*

Proof. Recall that we only work on some finite time interval $[0, T]$. The process (θ_t) is a Wiener process perturbed by some absolutely continuous change of probability measure: $d\mathbb{Q} = D_T d\mathbb{P}$. From the equation (37), we see that this can be reformulated as follows: The law of the process (θ_t) is absolutely continuous with respect to the Wiener measure, with density $w \mapsto \mathrm{e}^{\phi(T,w_T)}$. Now set $\eta_t = tX + W_t$ for every $t \leq T$. Conditionally on the vector X, the process (η_t) is just a Brownian motion plus a constant speed deterministic drift. As a result its law is explicit, given by a very basic version of the Cameron-Martin formula, see for instance [76, section 40]. For any test function H we have

$$\mathbb{E}(H(\eta) \mid X) = \mathbb{E}\left(H(W) \cdot \mathrm{e}^{X \cdot W_T - \frac{T}{2}|X|^2} \mid X\right).$$

Taking expectation again, and using Fubini and the definition (38) of ϕ, we obtain

$$\mathbb{E}H(\eta) = \mathbb{E}H(W) \cdot \mathrm{e}^{\phi(T,W_T)}.$$

Therefore, the law of (η_t) also has density $w \mapsto \mathrm{e}^{\phi(T,w_T)}$ with respect to the Wiener measure. □

Let us illustrate this result with a simple example where we can compute everything explicitly.

Example 50. In dimension 1, take μ to be the standard Gaussian measure. In that case we have an explicit formula for ϕ namely

$$\phi(t, \theta) = \frac{\theta^2}{2(1+t)} - \frac{1}{2}\log(1+t),$$

which gives $\nabla\phi(t, \theta) = \frac{\theta}{1+t}$. The equation for the tilt process (θ_t) is thus

$$d\theta_t = dW_t + \frac{\theta_t}{1+t}\, dt,$$

which can be solved explicitly:

$$\theta_t = (1+t)\int_0^t \frac{dW_s}{1+s}.$$

According to our theorem this should have the same law as the process (η_t) given by $\eta_t = W_t + tX$, where X is a standard Gaussian variable independent of (W_t). Of course this can be checked directly in this case. Indeed, both processes clearly are centered Gaussian processes and the two covariance structures coincide, since

$$\mathbb{E}\theta_s\theta_t = \mathbb{E}\eta_s\eta_t = st + s \wedge t.$$

for every $s, t > 0$. We leave this computation as an exercise.

6.3 Time reversal

We will now clarify the link between the stochastic localization of Eldan and the Gaussian localization of the previous section. Recall the definition (34) of μ_t. Letting ρ be the density of μ with respect to the Lebesgue measure we can reformulate this definition as

$$\int_{\mathbb{R}^n} f \, d\mu_t = \frac{\int_{\mathbb{R}^n} f(x)\rho(x) \exp\left(\theta_t \cdot x - t|x|^2/2\right) \, dx}{\int_{\mathbb{R}^n} \rho(x) \exp\left(\theta_t \cdot x - t|x|^2/2\right) \, dx}, \tag{40}$$

for any test function f. Let us introduce the heat semi-group

$$P_t f(x) = \mathbb{E} f(x + B_t) = f * g_t$$

where $g_t(x) = (2\pi t)^{-n/2} \mathrm{e}^{-|x|^2/2t}$ is the density of the Gaussian measure with mean 0 and covariance matrix $t \cdot \mathrm{Id}$. Warning: from now on (P_t) will denote the heat semigroup, and not the Langevin semigroup associated to μ from section 2.2. Then (40) rewrites as

$$\int_{\mathbb{R}^n} f \, d\mu_t = \frac{P_{1/t}(f\rho)}{P_{1/t}\rho} \left(\frac{\theta_t}{t} \right).$$

Now set $s = 1/t$. By Proposition 49 we have the following equality in law

$$\frac{\theta_t}{t} = \frac{tX + B_t}{t} = X + sB_{1/s}.$$

Since $\widetilde{B}_s := sB_{1/s}$ is again a standard Brownian motion (this is the time reversal property of the Brownian motion) we obtain the following: Up to the time reversal $t = 1/s$, the process $(\int f \, d\mu_t)_{t \geq 0}$ has the same distribution as $(Q_s f(X + B_s))_{s \geq 0}$, where Q_s is the operator defined by

$$Q_s f = \frac{P_s(f\rho)}{P_s\rho}.$$

Moreover, using the fact that the heat semigroup is self-adjoint in $L^2(dx)$ it is easy to see that

$$Q_s f(X + B_s) = \mathbb{E}[f(X) \mid X + B_s].$$

Putting everything together we see that the stochastic localization process (μ_t) initiated from μ has the same law as the measure-valued process obtained by looking at the conditional law of X given $X + B_s$ and then reversing time by setting $t = 1/s$. In particular if we take a snapshot at some fixed time $s = 1/t$, then for every test function f the variable $\int_{\mathbb{R}^n} f \, d\mu_t$ has the same law as $\mathbb{E}(f(X) \mid X + \sqrt{s}G)$ where G is a standard Gaussian vector independent of X.

Remark 51. It is clear from this description that this process was looked at in many other contexts. Apparently it is an important tool in filtering theory, and it is also very much related to what Bauerschmidt, Bodineau and Dagallier [10] call the Polcinski equation, which is used in their recent series of works on log-Sobolev inequalities for various particles systems.

7 Estimates for the conditional covariance

Our main task fin this section is to prove Theorem 41, which we reformulate here for convenience.

Theorem 52. *Let X be log-concave and isotropic, i.e. $\mathbb{E}X = 0$ and $\mathrm{Cov}(X) = \mathrm{Id}$, and let G be a standard Gaussian vector independent of X, then*

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \lesssim 1$$

for every s such that $(\log n)^2 \lesssim s$.

Recall that $\lesssim$ means up to a universal multiplicative constant (here a factor 10 is probably OK). Also the norm is the operator norm, which is also the maximal eigenvalue.

We shall derive this by combining arguments from Eldan [35], Lee-Vempala [63], Chen [32], Klartag-Lehec [55], with the improved Lichnerowicz inequality from Section 4. Actually the improved Lichnerowicz allows to bypass many ideas of the aforementioned papers.

Recall also that we have seen in Section 5 that the improved Lichnerowicz allows to show that if X is log-concave and if

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s} \cdot G)\|_{op} \lesssim 1,$$

for all $s \geq s_0$ then $C_P(X) \lesssim \sqrt{s_0}$. So the theorem indeed yields

$$C_P(X) \lesssim \log n,$$

which is the current best bound for the Poincaré constant of an isotropic log-concave random vector.

To prove the theorem we will reverse time and rewrite everything in terms of the stochastic localization process (μ_t) associated to μ. We thus rephrase Theorem 52 as follows.

Theorem 53. *Suppose μ is a log-concave and isotropic probability measure on $\mathbb{R}^n$ and let (μ_t) be the stochastic localization process initiated at μ. Then*

$$\mathbb{E}\|\mathrm{Cov}(\mu_t)\|_{op} \lesssim 1,$$

for all $t \leq c \cdot (\log n)^{-2}$, where $c > 0$ is a universal constant.

The point of this time reversal is that we can now control everything using Itô's formula and some convexity inequalities. The proof of the theorem requires some preliminaries. There will be a number of them, but taken individually, each of these is pretty easy.

7.1 The equation for the covariance

As we have seen in the previous section, for any test function f the martingale $M_t = \int_{\mathbb{R}^n} f\, d\mu_t$ satisfies

$$dM_t = \left(\int_{\mathbb{R}^n} f(x)(x - a_t)\, d\mu_t \right) \cdot dW_t,$$

where (W_t) is some standard Brownian motion. This obviously extends to vector valued functions. If $F \colon \mathbb{R}^n \to \mathbb{R}^k$ is a vector valued function that grows fairly reasonably at infinity then the process (M_t) given by

$$M_t = \int_{\mathbb{R}^n} F\, d\mu_t$$

is a martingale, and moreover

$$dM_t = \left(\int_{\mathbb{R}^n} F(x) \otimes (x - a_t)\, d\mu_t \right) \cdot dW_t$$

A bit more explicitly, writing x_i for the i-th coordinate of a vector $x \in \mathbb{R}^n$ we have

$$dM_t = \sum_{i=1}^{n} \left(\int_{\mathbb{R}^n} F(x)(x - a_t)_i\, d\mu_t \right) dW_{t,i}. \tag{41}$$

Lemma 54. *Let a_t and A_t be the barycenter and covariance matrix of μ_t, respectively. Then*

$$da_t = A_t dW_t$$
$$dA_t = \sum_{i=1}^{n} \left(\int_{\mathbb{R}^n} (x - a_t)^{\otimes 2}\, (x - a_t)_i\, d\mu_t \right) dW_{t,i} - A_t^2\, dt.$$

This is obtained by applying (41) to the tensors $F(x) = x$ and $F(x) = x \otimes x$ and then rearranging the terms appropriately. The details are left as an exercise.
This shows that the stochastic localization process has some moment generating property. The derivative for the barycenter is expressed in terms of the covariance, and the derivative for the covariance depends on 3-tensors.

7.2 Some matrix inequalities

Lemma 55. *Suppose K, H are symmetric matrices, and K is positive semi-definite. Then for every positive α, β we have*

$$\operatorname{Tr}(K^{\alpha} H K^{\beta} H) \leq \operatorname{Tr}(K^{\alpha+\beta} H^2).$$

Proof. Let $K = \sum \lambda_i x_i \otimes x_i$ be the spectral decomposition of K. Then

$$\begin{aligned}
\operatorname{Tr}(K^\alpha H K^\beta H) &= \sum_{ij} \lambda_i^\alpha \lambda_j^\beta \langle H x_i, x_j \rangle^2 \\
&\leq \sum_{ij} \lambda_i^{\alpha+\beta} \langle H x_i, x_j \rangle^2 \\
&= \sum_{i} \lambda_i^{\alpha+\beta} |H x_i|^2 \\
&= \sum_{i} \lambda_i^{\alpha+\beta} \langle H^2 x_i, x_i \rangle = \operatorname{Tr}(K^{\alpha+\beta} H^2).
\end{aligned}$$

The only inequality in the above display follows from Young's inequality

$$\lambda_i^\alpha \lambda_j^\beta \leq \frac{\alpha}{\alpha+\beta} \lambda_i^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} \lambda_j^{\alpha+\beta},$$

and the fact that the expression $\langle H x_i, x_j \rangle^2$ is symmetric in i and j. □

Corollary 56. *Let ϕ be the map defined on the space $S_n(\mathbb{R})$ of symmetric matrices by $\phi(A) = \operatorname{Tr} \mathrm{e}^A$. Then for every symmetric matrices A, H we have*

$$\nabla^2 \phi(A)(H, H) \leq \nabla \phi(A) \cdot H^2 = \operatorname{Tr}(\mathrm{e}^A H^2),$$

where $\nabla^2 \phi(A)$ stands for the Hessian matrix at A, viewed as a bilinear form on $S_n(\mathbb{R})$.

Proof. Assume first that the matrix A is positive. Then by the previous lemma we have

$$\begin{aligned}
\nabla^2 \phi(A)(H, H) &= \sum_{k \geq 1} \frac{1}{k!} \sum_{l=0}^{k-1} \operatorname{Tr}(A^l H A^{k-1-l} H) \\
&\leq \sum_{k \geq 1} \frac{1}{k!} \cdot k \cdot \operatorname{Tr}(A^{k-1} H^2) = \operatorname{Tr}(\mathrm{e}^A H^2),
\end{aligned}$$

which is the desired inequality. This argument does not work if A has some negative eigenvalues, but observe that the function ϕ has the property that

$$\phi(A + t \cdot \mathrm{Id}) = \mathrm{e}^t \phi(A)$$

By differentiating this equality with respect to A we see also $\nabla \phi$ and $\nabla^2 \phi$ satisfy the same equation, which means that adding a multiple of identity to A does not perturb the desired inequality. Therefore it is enough to prove it for positive A. □

7.3 Inequalities for 3-tensors

Recall the equation for A_t

$$dA_t = \sum_{i=1}^{n} H_{i,t} dW_i - A_t^2\, dt,$$

where

$$H_{i,t} = \int_{\mathbb{R}^n} (x - a_t)^{\otimes 2} (x - a_t)_i \, d\mu_t.$$

Recall that a_t is the barycenter of μ_t. So the matrix $H_{i,t}$ is of the form $\mathbb{E} X_i X^{\otimes 2}$ for some random vector with mean 0. We need to control such quantities. This is the purpose of the next two lemmas.

Lemma 57. *Let X be a centered log-concave vector. Then*

$$\sup_{u \in \mathbb{S}^{n-1}} \{ \| \mathbb{E}(X \cdot u) X^{\otimes 2} \|_{op} \} \lesssim \| \mathrm{Cov}(X) \|_{op}^{3/2}.$$

Proof. Let u, v be unit vector and let $H_u = \mathbb{E}(X \cdot u) X^{\otimes 2}$. By Cauchy-Scwharz

$$H_u v \cdot v = \mathbb{E}(X \cdot u)(X \cdot v)^2 \leq (\mathbb{E}(X \cdot u)^2)^{1/2} (\mathbb{E}(X \cdot v)^4)^{1/2}.$$

Now we use log-concavity. The variable $X \cdot v$ is a log-concave random variable centered at 0. We saw in the first section that moments of 1D log-concave measures satisfy a reverse Hölder inequality. In particular the fourth moment and second moment squared are of the same order. We thus get

$$H_u v \cdot v \leq C (\mathbb{E}(X \cdot u)^2)^{1/2} \mathbb{E}(X \cdot v)^2 \leq C \| \mathrm{Cov}(X) \|_{op}^{3/2}.$$

Taking the supremum in both u and v yields the result. □

Lemma 58. *Let X be a centered random vector satisfying the Poincaré inequality. Then*

$$\| \sum_{i=1}^{n} (\mathbb{E} X_i X^{\otimes 2})^2 \|_{op} \leq 4 C_P(X) \cdot \| \mathrm{Cov}(X) \|_{op}^2.$$

Proof. Recall the definition of H_u. When u is a coordinate vector e_i we write H_i rather than H_{e_i}. We need to show that for every unit vector u

$$\sum_{i=1}^{n} H_i^2 u \cdot u \leq 4 C_P(X) \cdot \| \mathrm{Cov}(X) \|_{op}^2.$$

An elementary computation shows that $\sum H_i^2 u \cdot u = \mathrm{Tr}(H_u^2)$. Moreover, since X is centered, we get from Cauchy-Schwarz and the Poincaré inequality

$$\begin{aligned}
\mathrm{Tr} H_u^2 &= \mathbb{E}(X \cdot u)(H_u X \cdot X) \\
&\leq (\mathbb{E}(X \cdot u)^2)^{1/2} \cdot (\mathrm{Var}(H_u X \cdot X))^{1/2} \\
&\leq (\mathbb{E}(X \cdot u)^2)^{1/2} \cdot (4 C_P(X) \mathbb{E} |H_u X|^2)^{1/2} \\
&= (\mathrm{Cov}(X) u \cdot u)^{1/2} \cdot (4 C_P(X) \mathrm{Tr}(H_u^2 \mathrm{Cov}(X)))^{1/2} \\
&\leq \| \mathrm{Cov}(X) \|_{op} \cdot (4 C_P(X) \mathrm{Tr}(H_u^2))^{1/2}.
\end{aligned}$$

Thus $\mathrm{Tr} H_u^2 \leq 4 C_P(X) \|\mathrm{Cov}(X)\|_{op}^2$, which is the result. ☐

Remark 59. We only applied Poincaré to a quadratic form so in a sense we only need a weak notion of Poincaré here. This observation will not be needed in the subsequent analysis but it was crucial in the original work of Eldan.

7.4 Freedman's inequality

Lastly we need a relatively classical deviation inequality for martingales, which is usually attributed to Freedman [39].

Lemma 60. *Let $(M_t)_{t\geq 0}$ be a continuous local martingale satisfying $M_0 = 0$. Then for every positive u and σ^2 we have*

$$\mathbb{P}(\exists t > 0 \colon M_t \geq u \ \textit{ and } \ \langle M \rangle_t \leq \sigma^2) \leq \mathrm{e}^{-u^2/2\sigma^2}.$$

Proof. We only sketch the argument and leave the details as an exercise. Start by proving the following statement: If (Z_t) is a square integrable martingale satisfying $\langle Z \rangle_t \leq \sigma^2$ for all $t > 0$ and almost surely, then $Z_\infty = \lim_{t\to+\infty} Z_t$ exists and satisfies

$$\mathbb{P}(Z_\infty \geq u) \leq \mathrm{e}^{-u^2/2\sigma^2}$$

for all $u > 0$. Coming back to Freedman's inequality, introduce the stopping time

$$\tau = \inf\{t > 0 \colon \langle M \rangle_t > \sigma^2\}$$

and apply the above statement to the martingale (M_t) stopped at time τ. ☐

7.5 The bound on the covariance matrix

Theorem 61. *Suppose μ is log-concave and isotropic on $\mathbb{R}^n$, and let (A_t) be the covariance process of the stochastic localization associated to μ. Then*

$$\mathbb{P}\left(\exists s \leq t \colon \|A_s\|_{op} \geq 2\right) \leq \exp\left(-\frac{1}{Ct}\right), \qquad \forall t \leq \frac{1}{C \log^2 n}.$$

Remark 62. We will see later on that this bound is pretty much sharp.

Proof. A common method to control the norm of a symmetric random matrix A is to use the Schatten norm $(\mathrm{Tr} A^p)^{1/p}$ where p is an even integer of order $\log n$ as a proxy for $\|A\|_{op}$. This is what Eldan does in his 2014 paper. For some reason we prefer to use another proxy, namely

$$h_\beta(M) := \frac{1}{\beta} \log \mathrm{Tr}\, \mathrm{e}^{\beta M}.$$

Note that h_β is a smooth function. Also

$$\lambda_{\max}(M) \leq \frac{1}{\beta} \log \mathrm{Tr}\, \mathrm{e}^{\beta M} \leq \lambda_{\max}(M) + \frac{\log n}{\beta}.$$

Therefore if β is of order $\log n$ then $h_\beta(M)$ is approximately the same as the maximal eigenvalue of M, up to an additive constant. Recall the equation for (A_t). From Itô's formula we get (omitting the time variable)

$$dh_\beta(A) = \nabla h_\beta(A) \cdot \sum_{i=1}^n H_i dB_i - \nabla h_\beta(A) \cdot A^2 \, dt + \frac{1}{2} \sum_{i=1}^n \nabla^2 h_\beta(A)(H_i, H_i) \, dt.$$

Let

$$M = \nabla h_\beta(A) = \frac{\mathrm{e}^{\beta A}}{\mathrm{Tr}(\mathrm{e}^{\beta A})},$$

and note that this is a positive semi-definite matrix of trace 1. Using Corollary 56, we see that the second derivative of h_β satisfies

$$\nabla^2 h_\beta(A)(H_i, H_i) \leq \beta \mathrm{Tr}(M H_i^2).$$

Dropping some negative terms we finally arrive at

$$dh_\beta(A) \leq \sum_{i=1}^n \mathrm{Tr}(M H_i) dB_i + \frac{\beta}{2} \mathrm{Tr} \left(M \sum_{i=1}^n H_i^2 \right) dt.$$

Let us deal with the absolutely continuous part. Since M is positive and has trace 1, we get from Lemma 58

$$\mathrm{Tr} \left(M \sum_{i=1}^n H_i^2 \right) \leq \left\| \sum_{i=1}^n H_i^2 \right\|_{op} \leq 4 C_P(\mu_t) \|A_t\|_{op}^2.$$

Recall that (μ_t) gets more and more log-concave along time. In particular if the original measure μ is log-concave then μ_t is t-uniformly log-concave, almost surely. From the improved Lichnerowicz inequality, Theorem 33, we get

$$C_P(\mu_t) \leq \left(\frac{\|A_t\|_{op}}{t} \right)^{1/2},$$

hence

$$dh_\beta(A) \leq \sum_{i=1}^n \mathrm{Tr}(M H_i) dB_i + \frac{2\beta}{\sqrt{t}} \cdot \|A_t\|_{op}^{5/2} dt.$$

Let us now bound the quadratic variation of the martingale part. For any unit vector u, letting $H_u = \sum H_i u_i$ we get from Lemma 57

$$\sum_{i=1}^n \mathrm{Tr}(M H_i) u_i = \mathrm{Tr}(M H_u) \leq \|H_u\|_{op} \leq C_0 \|A_t\|_{op}^{3/2}.$$

Therefore,

$$\sum_{i=1}^n \mathrm{Tr}(M H_i)^2 \leq C_0^2 \|A_t\|_{op}^3.$$

Let us summarize what we have obtained so far:

$$\begin{aligned} \|A_t\|_{op} \le h_\beta(A_t) &\le h_\beta(A_0) + Z_t + 2\beta \int_0^t s^{-1/2} \|A_s\|_{op}^{5/2}\, ds \\ &= 1 + \frac{\log n}{\beta} + Z_t + 2\beta \int_0^t s^{-1/2} \|A_s\|_{op}^{5/2}\, ds \end{aligned} \tag{42}$$

where (Z_t) is a continuous martingale starting from 0 whose quadratic variation satisfies

$$\langle Z \rangle_t \le C_1 \int_0^t \|A_s\|_{op}^3\, ds. \tag{43}$$

Now choose $\beta = 2\log n$, and assume that there exists $s \le t$ such that $\|A_s\|_{op} \ge 2$. If s is the smallest such time then before time s the operator norm of A is less than 2, so by (42)

$$2 = \|A_s\|_{op} \le \frac{3}{2} + Z_s + C_2 s^{1/2} \log n \le \frac{3}{2} + Z_s + C_2 t^{1/2} \log n$$

where C_2 is some constant. If t is a sufficiently small multiple of $(\log n)^{-2}$ then the latest inequality implies that $Z_s \ge \frac{1}{4}$. Moreover, thanks to (43) we also have $\langle Z \rangle_s \le C_3 s \le C_3 t$. Therefore,

$$\mathbb{P}(\exists s \le t\colon \|A_s\|_{op} \ge 2) \le \mathbb{P}(\exists s > 0\colon Z_s \ge \frac{1}{4} \text{ and } \langle Z \rangle_s \le C_3 t).$$

We conclude with Freedmann's inequality, Lemma 60. □

Now we prove the bound for the expectation of A_t.

Proof of Theorem 53. Since μ_t is t-uniformly log-concave, its covariance matrix is bounded above by $(1/t)\mathrm{Id}$. This was already mentioned in Section 4. Therefore we have $\|A_t\|_{op} \le 1/t$, almost surely. As a result

$$\mathbb{E}\|A_t\|_{op} \le 2 + \frac{1}{t}\mathbb{P}(\|A_t\|_{op} > 2).$$

Now we apply the latest theorem. Since $x \cdot \mathrm{e}^{-c_1 x}$ is a bounded function of x we indeed get $\mathbb{E}\|A_t\|_{op} \lesssim 1$ on the time range $[0, (C\log n)^{-2}]$. □

Remark 63. Instead of the improved Lichnerowicz inequality, we could have bounded $C_P(\mu_t)$ by the KLS constant. Namely if C_n is the largest Poincaré constant of an isotropic log-concave measure then it is easy to see that for any log-concave X

$$C_P(X) \le C_n \|\mathrm{Cov}(X)\|_{op}.$$

Therefore

$$C_P(\mu_t) \le C_n \|A_t\|_{op}.$$

Using this estimate instead of the improved Lichnerowicz inequality leads to the following statement:

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \lesssim 1, \quad \text{provided } C_n \log n \lesssim s. \tag{44}$$

This is also good enough for the $\log n$ bound for C_n. Indeed we have seen that if the expected norm of $\mathrm{Cov}(X \mid X + \sqrt{s}G)$ is of order 1 for all $s \geq s_0$ then $C_P(\mu) \lesssim \sqrt{s_0}$. So the latest display actually gives $C_n \lesssim \sqrt{C_n \log n}$, hence $C_n \lesssim \log n$.

Remark 64. We will see later on an example of a measure for which $\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op}$ explodes at times $s = \log n$ (but is bounded at time $10 \log n$). In a sense this is evidence for the KLS conjecture $C_n \lesssim 1$ to be indeed correct. Namely if KLS is correct then (44) is sharp and the above analysis of the conditional covariance is essentially the best one can do.

7.6 Life before improved Lichnerowicz

The improved Lichnerowicz estimate is only from 2023, and it was not available to Eldan, Lee-Vempala, Chen, Klartag-Lehec. Still these authors gave non trivial estimate on the KLS constant using this localization technique. In particular the KL bound was polynomial in $\log n$. Here we will only say a few words about the original argument of Eldan.

First, let us derive a bound on the Poincaré constant of μ from a bound on the covariance of the stochastic localization in a slightly different manner than what was done in the previous section. Let f be the function given by E. Milman's result (see section 2.3). Namely f is 1-Lipschitz and such that

$$\mathrm{Var}_\mu(f) \approx \|f\|_\infty^2 \approx C_P(\mu).$$

By the decomposition of variance

$$\mathrm{Var}_\mu(f) = \mathbb{E}\mathrm{Var}_{\mu_t}(f) + \mathrm{Var}\left(\int_{\mathbb{R}^n} f \, d\mu_t\right)$$

For the first term we proceed in the same way as before: by improved Lichnerowicz and since f is 1-Lipschitz, we have

$$\mathbb{E}\mathrm{Var}_{\mu_t}(f) \leq \frac{\mathbb{E}\|A_t\|_{op}^{1/2}}{\sqrt{t}}.$$

For the second term, we proceed differently. The process $M_t = \int f \, d\mu_t$ is a martingale, whose derivative is

$$dM_t = \left(\int_{\mathbb{R}^n} f(x)(x - a_t) \, d\mu_t\right) \cdot dW_t.$$

Since $\|f\|_\infty^2 \lesssim C_P(\mu)$ we get from Cauchy-Schwarz

$$\left|\int_{\mathbb{R}^n} f(x)(x - a_t) \, d\mu_t\right|^2 \lesssim C_P(\mu) \|A_t\|_{op}.$$

Hence

$$\mathrm{Var}\left(\int f \, d\mu_t\right) \lesssim C_P(\mu) \int_0^t \mathbb{E}\|A_s\|_{op} \, ds.$$

If $\mathbb{E}\|A_t\|_{op} \lesssim 1$ up until time $t_0 \lesssim 1$ we finally get

$$C_P(\mu) \lesssim t^{-1/2} + t \cdot C_P(\mu),$$

for all $t \leq t_0$, which indeed implies $C_P(\mu) \lesssim t_0^{-1/2}$. One thing that we can notice from this proof is that if we replace the improved Lichnerowicz inequality by the usual one, namely $C_P(\mu) \leq 1/t$ in the t-uniformly log-concave case, we also get something non trivial, namely

$$C_P(\mu) \lesssim t_0^{-1}. \tag{45}$$

This is obviously a lot worse than what we get from improved Lichnerowicz, but still non trivial. As a matter of fact, all the aforementioned works on the KLS conjecture (prior to the latest one by Klartag in which the improved Lichnerowicz inequality is established) rely on this estimate, one way or another. The other argument to get Poincaré from the bound on the conditional covariance (see section 5) may be more elementary and more natural in a way, but it only works if one happens to know the improved Lichnerowicz inequality. If you combine it with the usual Lichnerowicz inequality you get nothing.
Now we define the constants K_n and S_n by

$$K_n = \sup \left\{ \left\| \sum_{i=1}^n (\mathbb{E} X_i X^{\otimes 2})^2 \right\|_{op} \right\}, \quad S_n = \sup \left\{ \frac{1}{n} \mathrm{Var} |X|^2 \right\}$$

where both sup are taken over all log-concave isotropic random vectors on $\mathbb{R}^n$. The constant S_n is called the thin-shell constant. The thin-shell conjecture asserts that the sequence (S_n) is bounded. This is a weak form of the KLS conjecture as we only require Poincaré for a very specific function, namely the Euclidean norm squared. It was mentioned in the first section in connection with the central limit problem for convex sets. A variant of what we have done above shows that in the isotropic log-concave case we have $\mathbb{E}\|A_t\|_{op} \lesssim 1$ up until times $(CK_n \log n)^{-1}$. From (45) we then obtain the following bound

$$C_n \lesssim K_n \log n,$$

for the KLS constant C_n. Moreover, by definition of S_n, given a log-concave and isotropic vector X on $\mathbb{R}^n$, a unit vector u, and an orthogonal projection P of rank k, we have $\mathbb{E}(X \cdot u)|PX|^2 \leq \sqrt{kS_k}$. Applying this to suitable chosen projections P one can estimate the eigenvalues of $\mathbb{E}(X \cdot u)X^{\otimes 2}$ and then arrive at the bound

$$K_n \lesssim \sum_{k=1}^n \frac{S_k}{k} \lesssim S_n \log n.$$

We refer to [35] for the details. Altogether this gives

$$C_n \lesssim S_n (\log n)^2. \tag{46}$$

In other words thin-shell implies KLS up to polylog. This was the original result of Eldan. Note that Exercise 8 implies in particular that $S_n \geq 2$ for all n. As a result (46) has become irrelevant now that we know that $C_n \lesssim \log n$.

8 Further localization results

8.1 An obstruction to a full solution of KLS

As we have seen above, the KLS conjecture would be implied by the following statement: in the isotropic log-concave case the expected operator norm of $\mathrm{Cov}(X \mid X + \sqrt{s}G)$ remains of order 1 for all s. Unfortunately such an estimate cannot be true as we shall see now.

Let $X = (X_1, \dots, X_n)$ be a random vector whose coordinates are i.i.d. and such that $1 + X_i$ is an exponential variable of parameter 1. This is clearly an isotropic log-concave vector on $\mathbb{R}^n$.

Proposition 65. *We have $\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \leq C$ for all $s \geq C \log n$. On the other hand, if $s \leq \log n$ then $\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \geq cs$.*

Note that from the tensorization property of the Poincaré inequality (see exercise 1), we have $C_P(X) = C_P(X_1)$. In particular the Poincaré constant of X does not depend on n and X is not a counterexample to the KLS conjecture. Recall also that we always have the bound

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \leq s$$

(acually this is true almost surely, not only in expectation). This example shows that this bound can be essentially sharp on a time range $[0, s_0]$ with $s_0 \to \infty$, namely $s_0 = \log n$. In particular at time s_0 we have

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s_0}G)\|_{op} \geq c \log n$$

so the expected norm of the conditional covariance is not bounded for all times. In view of this example, the best one could hope for is

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \lesssim 1, \quad \forall s \geq C \log n \tag{47}$$

and for every log-concave isotropic X. Notice that there is still a gap between this and the bound that we obtained in our theorem (in which $\log n$ is replaced by $\log^2 n$). If true the estimate (47) would imply the bound

$$C_n \lesssim (\log n)^{1/4},$$

for the KLS constant. This seems to be the limit one could reach within this framework. Going below this mark would have to rely on different arguments.

The proof of the proposition only relies on some analysis in one dimension. Indeed since the coordinates of X and G are all independent it is clear that the conditional law of X given $X + \sqrt{s}G$ is just the n-fold product of the law of X_1 condition on $X_1 + \sqrt{s}G_1$. So the conditional covariance is diagonal with i.i.d. entries, and we just have to estimate the expected maximum of these.

As a preliminary step, we need to compute the variance of a truncated Gaussian.

Lemma 66. *Let g be as standard Gaussian variable, then*

$$\mathrm{Var}(g \mid g \geq x) \approx \frac{1}{1 + x_+^2}.$$

Proof. Observe that the conditional law of g given $g \geq x$ is log-concave. Let us use the "How to think about 1D log-concave measures" proposition from the first section. It implies in particular that for log-concave random variable X having density f we have

$$\|f\|_\infty^2 \cdot \mathrm{Var}(X) \approx 1,$$

Applying this to the conditional law of g given $g \geq x$ (which is indeed log-concave), we get $\mathrm{Var}(g \mid g \geq x) \approx \left(\int_x^\infty \mathrm{e}^{-y^2/2}\, dy\right)^{-2}$ if $x \leq 0$ and

$$\mathrm{Var}(g \mid g \geq x) \approx \left(\mathrm{e}^{x^2/2} \int_x^\infty \mathrm{e}^{-y^2/2}\, dy\right)^{-2}$$

for every positive x. The result follows easily. □

Proof of Proposition 65. The conditional law of X_1 given $X_1 + \sqrt{s}G_1$ is just a truncated Gaussian. After some elementary computation we get

$$\mathrm{Var}(X_1 \mid X_1 + \sqrt{s}G_1) = s \cdot v(\sqrt{s} - \frac{1}{\sqrt{s}}Y_1 - G_1) \tag{48}$$

where $Y_1 = X_1 + 1$ and v is the function given by

$$v(x) = \mathrm{Var}(G_1 \mid G_1 \geq x).$$

Note that Y_1 is an exponential variable independent of G_1, hence

$$\mathbb{P}(Y_1 \geq s, G_1 \geq 0) = \frac{1}{2}\mathrm{e}^{-s}.$$

Since the function v is bounded away from 0 on $\mathbb{R}_-$, if $Y_1 \geq s$ and $G_1 \geq 0$ then

$$\mathrm{Var}(X_1 \mid X_1 + \sqrt{s}G_1) \geq cs.$$

As a result

$$\mathbb{P}(\mathrm{Var}(X_1 \mid X_1 + \sqrt{s}G_1) \geq cs) \geq \frac{1}{2}\mathrm{e}^{-s}.$$

By independence we get

$$\mathbb{P}\left(\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \geq cs\right) \geq 1 - (1 - \frac{1}{2}\mathrm{e}^{-s})^n$$

If $s \leq \log n$, the right-hand side is at least $1 - \mathrm{e}^{-1/2}$. By Markov's inequality, this implies that

$$\mathbb{E}\|\mathrm{Cov}(X \mid X + \sqrt{s}G)\|_{op} \geq c'.$$

For the other inequality, since $v(x) \lesssim x^{-2}$ for large x, equation (48) and the union bound imply in particular that if C is a sufficiently large constant

$$\mathbb{P}(|\mathrm{Var}(X_1 \mid X_1 + \sqrt{s}G_1)| \geq C) \leq \mathbb{P}(Y_1 \geq \frac{s}{4}) + \mathbb{P}(G_1 \geq \frac{\sqrt{s}}{4}) \leq 2\mathrm{e}^{-cs}.$$

Hence, by the union bound again,

$$\mathbb{P}(\|\mathrm{Cov}(X \mid X+\sqrt{s}G)\|_{op} \geq C) \leq 2n\mathrm{e}^{-cs}.$$

Since $\|\mathrm{Cov}(X \mid X+\sqrt{s}G)\|_{op} \leq s$ almost surely this implies

$$\mathbb{E}\|\mathrm{Cov}(X \mid X+\sqrt{s}G)\|_{op} \leq 2ns \cdot \mathrm{e}^{-cs} + C.$$

This becomes $O(1)$ as soon as s exceeds a sufficiently large multiple of $\log n$. □

8.2 Concentration of measure

Recall from section 2.1 the definition of the concentration function of μ:

$$\alpha_\mu(r) = \sup\left\{1-\mu(S_r) \colon \mu(S) \geq 1/2\right\}.$$

We saw that log-concave measures satisfy exponential concentration and moreover than the Poincaré constant and the exponential concentration constant squared are of the same order. In particular the current best estimate for KLS amounts to the following

$$\alpha_\mu(r) \leq 2\exp\left(-c \cdot \frac{r}{\sqrt{\log n}}\right). \tag{49}$$

One of the points of this section is to show that one can go a bit beyond this estimate.

Theorem 67. *If μ is log-concave and isotropic then its concentration function satisfies*

$$\alpha_\mu(r) \leq 2\exp\left(-c \cdot \min\left(r, \frac{r^2}{\log^2 n}\right)\right), \quad \forall r \geq 0. \tag{50}$$

We should make some comments on this result. First of all, the rate provided by Theorem 67 is not smaller than that of (49) on the whole halfline. In particular combining with E. Milman's theorem would only lead to a $(\log n)^2$ bound for the Poincaré constant of an isotropic log-concave measure (rather than $\log n$). That being said, the theorem yields in particular the rate e^{-cr}, which is predicted by the KLS conjecture, as soon as r is larger than $\log^2 n$ or so. As far as we know, this information cannot be inferred from the bound $C_n \lesssim \log n$ alone. Let us also mention that one can prove the following variant of (50), in which the concentration depends on the KLS constant C_n:

$$\alpha_\mu(r) \leq 2\exp\left(-c \cdot \min\left(r, \frac{r^2}{C_n \cdot \log n}\right)\right), \quad \forall r \geq 0. \tag{51}$$

This inequality is taken from Bizeul [13].

This concentration is reminiscent of the Guédon-Milman estimate from 2011, see [43]. They proved that every isotropic log-concave measure μ satisfies

$$\mu\left(\left||x| - \sqrt{n}\right| \geq r\right) \leq 2\exp\left(-c \cdot \min\left(r, \frac{r^2}{n^{2/3}}\right)\right), \quad \forall r \geq 0.$$

This is weaker than (50) in two ways, first of all the constant is much worse ($n^{2/3}$ vs $\log^2 n$) and the deviation inequality is only for the Euclidean norm, and not for every 1-Lipschitz function, as in (50). This application of stochastic localization to concentration dates back to Lee and Vempala [63]. Their main result in that paper is the bound $C_n = O(n^{1/2})$ for the KLS constant, but they also obtain the inequality

$$\alpha_\mu(r) \leq 2\exp\left(-c \cdot \min\left(r, \frac{r^2}{n^{1/2}}\right)\right). \tag{52}$$

In contrast with (51), they do not loose a logarithm when they pass from the bound on the KLS constant to the deviation inequality. Their argument is very delicate and clever but it only works with a polynomial estimate for C_n and it does not allow to remove the logarithm from (51) now that we have a logarithmic estimate for C_n.

The proof of Theorem 67 relies on the fact that uniformly log-concave measures satisfy Gaussian concentration. This was already mentioned in section 2.4.

Proposition 68. *Let μ be a t-uniformly log-concave measure. Then for every measurable set S and every $r \geq 0$ we have*

$$\mu(S)(1 - \mu(S_r)) \leq \exp(-c \cdot tr^2), \quad \forall r \geq 0,$$

where c is a universal constant. In particular the Gaussian concentration constant of μ is $O(t^{-1/2})$.

Proof. We give a short proof based on the Prékopa-Leindler inequality: if f, g, h are non negative functions on $\mathbb{R}^n$ satisfying the inequality

$$\sqrt{f(x)g(y)} \leq h\left(\frac{x+y}{2}\right)$$

for every $x, y \in \mathbb{R}^n$ then

$$\sqrt{\int_{\mathbb{R}^n} f(x)\,dx \int_{\mathbb{R}^n} g(x)\,dy} \leq \int_{\mathbb{R}^n} h(x)\,dx.$$

If μ is t-uniformly log-concave then its potential V satisfies

$$V\left(\frac{x+y}{2}\right) \leq \frac{V(x)+V(y)}{2} - \frac{t}{8}|x-y|^2.$$

Given a set S and $\theta > 0$, one can then see that the hypothesis of Prékopa-Leinder applies to the functions $f(x) = \mathbf{1}_S(x)\mathrm{e}^{-V(x)}$, $g(y) = \mathrm{e}^{\theta d(y,S) - V(y)}$ and $h = \mathrm{e}^{-V + 2\theta^2/t}$. From the conclusion of Prékopa, we get

$$\mu(S) \cdot \int_{\mathbb{R}^n} \mathrm{e}^{\theta d(x,S)}\,d\mu \leq \mathrm{e}^{2\theta^2/t}.$$

The conclusion then follows from Chernoff inequality. □

One can be a bit more precise. As we already mentioned, in the Gaussian case we know the exact value of the concentration function α_{γ_n}. Indeed, an integrated version of the isoperimetric inequality of Sudakov-Tsirelson / Borell asserts that $\gamma_n(S_r)$ is maximized when S is a halfspace. In particular

$$\alpha_{\gamma_n}(r) = 1 - \Phi(r), \quad \forall r \geq 0.$$

where Φ is the distribution function of the standard Gaussian variable. Moreover, a deep result of Caffarelli asserts that a measure μ that is more log-concave than a given Gaussian measure is the image of that Gaussian by a 1-Lipshitz map. Besides, it is clear that a pushforward by a contraction can only lower the concentration function. As a result if μ is t-uniformly log-concave then its concentration function is upper bounded by that of the Gaussian measure with covariance Id/t, namely we have

$$\alpha_\mu(r) \leq 1 - \Phi(\sqrt{t} \cdot r), \quad \forall r \geq 0.$$

Since $1 - \Phi(r) \leq \frac{1}{2}\mathrm{e}^{-r^2/2}$ for $r \geq 0$, this implies Gaussian concentration. However this only improves upon Proposition 68 at the level of the value of the universal constant c, which is irrelevant for our purposes.

Proof of Theorem 67. Fix a set S of measure $1/2$ and write

$$1 - \mu(S_r) = \mathbb{E}(1 - \mu_t(S_r)) \leq \mathbb{E}(1 - \mu_t(S_r))\mathbb{1}_{\{\mu_t(S) \geq 1/4\}} + \mathbb{P}(\mu_t(S) \leq 1/4),$$

where (μ_t) is the stochastic localization of μ. Since μ_t is t-uniformly log-concave, the first term is at most $4\mathrm{e}^{-ctr^2}$, by Proposition 68. To handle the second term recall that the martingale $M_t := \mu_t(S)$ satisfies

$$dM_t = \left(\int_S (x - a_t)\, d\mu_t \right) \cdot dW_t.$$

Applying Cauchy-Schwarz we obtain

$$\left| \int_S (x - a_t)\, d\mu_t \right|^2 \leq \mu_t(S) \|A_t\|_{op} \leq \|A_t\|_{op}.$$

Hence the inequality

$$\langle M \rangle_t \leq \int_0^t \|A_s\|_{op}\, ds.$$

In particular if $\|A_s\|_{op} \leq 2$ on $[0, t]$ then $\langle M \rangle_t \leq 2t$. Therefore

$$\mathbb{P}(M_t \leq \frac{1}{4}) \leq \mathbb{P}(M_t \leq \frac{1}{4} \text{ and } \langle M \rangle_t \leq 2t) + \mathbb{P}(\exists s \leq t \colon \|A_s\|_{op} \geq 2).$$

By Theorem 61 from the previous section, the second term is at most $\exp(-(Ct)^{-1})$, provided $t \leq (C \log^2 n)^{-1}$. On the other hand since $M_0 = \mu(S) = 1/2$, Freedman's inequality (Lemma 60) insures that

$$\mathbb{P}(M_t \leq \frac{1}{4} \text{ and } \langle M \rangle_t \leq 2t) \leq \exp\left(-\frac{1}{C_1 t} \right).$$

Putting everything together we get

$$\mu(S_r^c) \leq 4\exp(-c \cdot tr^2) + 2\exp(-(C_2 t)^{-1})$$

for every $t \leq (C \cdot \log n)^{-2}$. Choosing $t = \min(r^{-1}, (C \cdot \log n)^{-2})$ yields

$$\mu(S_r^c) \leq 6\exp\left(-c' \cdot \min\left(r, \frac{r^2}{\log^2 n}\right)\right).$$

One can replace the prefactor 6 by 2 by changing a bit the constant c' in the exponent. □

Here is an example of an application of the theorem.

Corollary 69 (Paouris theorem)**.** *Suppose μ is log-concave and isotropic then*

$$\mu(|x| \geq r) \leq \exp(-cr), \quad \forall r \geq C\sqrt{n},$$

where as usual c, C are universal constants.

The inequality is due to Paouris [72], see also [1] for another proof. The inequality can also be expressed in terms of moments. It asserts that if X is log-concave and isotropic on $\mathbb{R}^n$ then the moments of the Euclidean norm of X remain constant for quite a while, namely

$$(\mathbb{E}|X|^p)^{1/p} \approx (\mathbb{E}|X|^2)^{1/2}$$

for p as large as $\sqrt{n}$.

Proof. We apply the concentration estimate to the 1-Lipschitz function $f(x) = |x|$. We get in particular

$$\mu(|x| \geq m + r) \leq \mathrm{e}^{-cr},$$

provided that $r \geq C \cdot \log^2 n$, where m is a median for $|x|$. Since $m \leq 2\int |x|\, d\mu \leq 2\sqrt{n}$, the latest display is easily seen to imply the desired inequality. □

Remark 70. As is apparent from the proof, inequality (50) is an overkill for this application, and the argument would go through using (52) instead. As a matter of fact this application to the Paouris inequality is taken from Lee and Vempala's paper [63].

8.3 Logarithmic Sobolev inequality and a variant of the KLS conjecture

Recall from section 2.5 that the log-Sobolev constant of a probability measure on $\mathbb{R}^n$, denoted $C_{LS}(\mu)$ is the best constant in the inequality

$$D(\nu \mid \mu) \leq \frac{1}{2} C_{LS}(\mu) I(\nu \mid \mu)$$

where D and I denote the relative entropy and Fisher information, respectively. Once again, the uniformly log-concave case is well understood.

Theorem 71 (Bakry-Émery criterion [5]). *If μ is t-uniformly log-concave then $C_{LS}(\mu) \le t^{-1}$. The inequality is sharp, equality is attained for the Gaussian measure of covariance $t^{-1} \cdot \mathrm{Id}$.*

There are many ways to prove this inequality, see for instance [30] for an overview. Again we are interested in the log-concave case. However, in contrast with the Poincaré inequality, not every log-concave measure satisfies log-Sobolev. Indeed, recall that log-Sobolev implies Gaussian concentration, with explicit control of the constants, and that this can be reversed in the log-concave case: the log-Sobolev constant and the Gaussian concentration constant are actually of the same order for log-concave measures. To insure log-Sobolev, one has to impose another condition on top of log-concavity, such as having bounded support. The following result is due to Lee-Vempala [63].

Theorem 72. *Suppose μ is log-concave, isotropic, and supported on a set of diameter D. Then $C_{LS}(\mu) \lesssim D$.*

Let us remark that because of the equivalence between log-Sobolev and Gaussian concentration in the log-concave case, a log-concave measure supported on a set of diameter D trivially has $O(D^2)$ log-Sobolev constant. Since the diameter of the support of an isotropic measure is at least $\sqrt{n}$ the theorem improves greatly upon the trivial bound in the isotropic case. It should also be noted that for the uniform measure on the ℓ_1 ball rescaled to be isotropic, the diameter of the support and the log-Sobolev constant both are of order n.

Proof. This is actually an easy consequence of our concentration result Theorem 67. Indeed, the latter asserts that if μ is log concave and isotropic then

$$\alpha_\mu(r) \le 2 \exp\left(-c \cdot \min(r, \frac{r^2}{\log^2 n}) \right), \quad \forall r \ge 0.$$

On the other hand if μ is supported on a set of diameter D then trivially $\alpha_\mu(r) = 0$ if $r > D$. On the interval $[0, D]$ we have $r \le r^2/D$, and since $D \ge \sqrt{n} \ge \log^2 n$, we finally obtain

$$\alpha_\mu(r) \le 2 \exp\left(-c' \cdot \frac{r^2}{D} \right).$$

The Gaussian concentration constant is thus $O(D)$, which implies the desired inequality by E. Milman's result, Theorem 26. ☐

Let us try to relax the bounded support assumption. We know that log-Sobolev implies Gaussian concentration. In particular linear functions should have sub-Gaussian tails, at a rate controlled by the log-Sobolev constant. Let us be a bit more precise.

Definition 73. *Suppose f is a function having mean zero for μ. We denote by $\|f\|_{\psi_2(\mu)}$ the Orlicz norm of f associated to the Orlicz function $\mathrm{e}^{r^2} - 1$, namely the best constant C in the inequality*

$$\mu(|f| \ge r) \le 2 \cdot \exp\left(-\frac{r^2}{C^2} \right).$$

Remark 74. The usual definition of the Orlicz norm of f associated to an Orlicz function ϕ and a measure μ is the smallest constant C such that $\int \phi(|f|/C)\, d\mu \leq 1$. The above definition is slightly different but equivalent, up to universal constants. We chose this modified definition so that the connection with the notion of Gaussian concentration from section 2.4 becomes more apparent.

The discussion above shows that for any probability measure and any direction θ we have

$$\|x \cdot \theta\|_{\psi_2(\mu)}^2 \lesssim C_{LS}(\mu).$$

It is natural to conjecture that this inequality could be reversed in the log-concave case. This amounts to saying that the log-Sobolev constant is the largest ψ_2-norm squared of a linear function, very much like the KLS conjecture predicts that the Poincaré constant of a log-concave measure is up to a constant the largest L^2-norm squared of a linear function.

Definition 75 (Log-Sobolev version of KLS constant, Bizeul [12])**.** *Let D_n be the largest log-Sobolev constant of a log-concave measure for which $\|x \cdot \theta\|_{\psi_2(\mu)} \leq 1$ for every direction θ.*

Conjecture 76 (Log-Sobolev KLS conjecture, Bizeul [12])**.**

$$D_n = O(1).$$

It follows from some result of Bobkov [16] from 2007 that $D_n = O(n)$. Using stochastic localization, one can show the following.

Theorem 77 (Bizeul [12])**.**

$$D_n = O(n^{1/2}).$$

Proof. The idea is to combine Theorem 67 with a rather crude net argument. Again, by E. Milman's theorem it is enough to prove that if μ log-concave is ψ_2 with norm at most 1 in all directions, then its concentration function satisfies

$$\alpha_\mu(r) \leq C \mathrm{e}^{-cr^2/\sqrt{n}}. \tag{53}$$

Note that the ψ_2 norm is larger than the L^2 norm, maybe up to a constant. So the covariance of μ has operator norm $O(1)$. The concentration function of μ thus satisfies

$$\alpha_\mu(r) \leq 2 \exp\left(-c \cdot \min(r, \frac{r^2}{\log^2 n}) \right),$$

for every $r > 0$. Here there is a small gap which we can leave as an exercise: show that having an upper bound for the concentration function of every isotropic log-concave μ of the form $\alpha_\mu \leq \alpha_*$ implies that if μ is log-concave but not necessarily isotropic, then $\alpha_\mu(r) \leq \alpha_*(r/\sqrt{\|\mathrm{Cov}(\mu)\|_{op}})$. We thus get an estimate that is smaller than our target concentration if $r > cr^2/\sqrt{n}$, namely $r < C\sqrt{n}$. Therefore, it is enough to prove (53) when r is a sufficiently large multiple of $\sqrt{n}$. Moreover, by Markov's inequality we have

$$\mu(|x| \geq 2\sqrt{n}) \leq \frac{1}{4n} \int_{\mathbb{R}^n} |x|^2 \, d\mu \leq \frac{1}{4}.$$

So if S is a set of measure $1/2$ then S intersects the ball of radius $2\sqrt{n}$. If $r \geq 2\sqrt{n}$ this implies easily that $S_{2r} \supset \{|x| \leq r\}$, hence

$$\alpha_\mu(2r) \leq \mu(|x| > r).$$

So it is enough to prove that $\mu(|x| > r) \leq \mathrm{e}^{-cr^2/\sqrt{n}}$ for $r \geq C\sqrt{n}$. Now recall the ψ_2 hypothesis: For every direction θ and every r, we have

$$\mu\{|x \cdot \theta| > r\} \leq 2\mathrm{e}^{-r^2}. \tag{54}$$

It is well-known that there exists $1/2$-net of the unit sphere of cardinality 5^n at most. Let N be such a set. Since any element x in the sphere is at distance $1/2$ at most from a point of N we have

$$|x| \leq 2 \max_{\theta \in N} \{x \cdot \theta\},$$

for every $x \in \mathbb{R}^n$. Applying (54) to every θ in the net and the union bound we get

$$\mu(|x| > r) \leq 2 \cdot 5^n \mathrm{e}^{-r^2/4}.$$

If $r > C\sqrt{n}$ for a sufficiently large constant C, we deduce from this inequality

$$\mu(|x| > r) \leq \mathrm{e}^{-r^2/8}$$

which is even better than what we needed. □

Remark 78. This proof seems to have lots of slack. It does not seem like the concentration result (Theorem 67) nor the ψ_2 hypothesis were fully exploited. In particular, it should be noted that the argument would go through with the weaker concentration estimate from Lee and Vempala (49). Nevertheless, as far as the log-Sobolev version of the KLS conjecture is concerned this is the best result around, as of today.

9 Bourgain's slicing problem

Consider a centrally-symmetric convex body $K \subseteq \mathbb{R}^n$ (i.e. $K = -K$). The maximal function operator associated with K, defined for $f : \mathbb{R}^n \to \mathbb{R}$ via

$$M_K f(x) = \sup_{r>0} \int_K f(x+ry) \frac{dy}{Vol_n(K)}.$$

Bourgain [23] proved that $\|M_K\|_{L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)} \leq C$ for a universal constant $C > 0$. This led him to study on another question, seemingly innocent:

Question 79. *Let $n \geq 2$ and suppose that $K \subseteq \mathbb{R}^n$ is a convex body of volume one. Does there exist a hyperplane $H \subseteq \mathbb{R}^n$ such that*

$$Vol_{n-1}(K \cap H) > c \tag{55}$$

for a universal constant $c > 0$?

This question is still not completely answered, and in the last four decades it emerged as an "engine" for the development of the research direction discussed in these lectures. It is shown in [54] that the bound (55) holds true if we replace the universal constant c on the right-hand side by $c/\sqrt{\log n}$. This is the currently best known result in the general case.

Theorem 80 (Hensley [44], Fradelizi [38])**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body whose barycenter lies at the origin. Let X be a random vector distributed uniformly in K, and assume that $\mathrm{Cov}(X)$ is a scalar matrix. Then for any $\theta_1, \theta_2 \in S^{n-1}$,*

$$Vol_{n-1}(K \cap \theta_1^\perp) \leq C \cdot Vol_{n-1}(K \cap \theta_2^\perp)$$

where $C > 0$ is a universal constant. In fact, $C \leq \sqrt{6}$.

Proof. Let $\theta \in S^{n-1}$ and denote

$$\sigma = \sqrt{\mathbb{E}(X \cdot \theta)^2} = \sqrt{\mathrm{Cov}(X)\theta \cdot \theta},$$

which is independent of θ. Write

$$\rho_\theta(t) = \frac{Vol_{n-1}(K \cap (t\theta + \theta^\perp))}{Vol_n(K)},$$

the density of the random variable $X \cdot \theta$. By the Brunn-Minkowski inequality, ρ_θ is a log-concave probability density. The log-concave random variable $X \cdot \theta / \sigma$ has mean zero and variance one, and its density is $x \mapsto \sigma \rho_\theta(x\sigma)$. According to Proposition 4 above, for any $x \in \mathbb{R}$,

$$c' 1_{\{|x| \leq c''\}} \leq \sigma \rho_\theta(x\sigma) \leq C e^{-c|x|}$$

In particular, $c \leq \rho_\theta(0) \leq C$, for some universal constants $c, C > 0$. ☐

From this proof we may obtain a few more conclusions. First, that among all hyperplane sections parallel to a given hyperplane, the hyperplane section through the barycenter has the largest volume, up to a multiplicative universal constant. Second, that when $K \subseteq \mathbb{R}^n$ is a centered convex body of volume one, for any $\theta \in S^{n-1}$,

$$Vol_{n-1}(K \cap \theta^{\perp}) \cdot \sqrt{\mathbb{E}(X \cdot \theta)^2} \sim 1.$$

Here $\theta^{\perp} = \{x \in \mathbb{R}^n \, ; \, x \cdot \theta = 0\}$ and we abbreviate $A \sim B$ if $c \cdot A \leq B \leq C \cdot A$ for universal constants $c, C > 0$. This leads to the following conclusion:

Corollary 81. *Let $K \subseteq \mathbb{R}^n$ be a convex body of volume one and let X be a random vector distributed uniformly on K. Then,*

$$\sup_H Vol_{n-1}(K \cap H) \sim \frac{1}{\sqrt{\|\mathrm{Cov}(X)\|_{op}}},$$

where the supremum runs over all hyperplanes $H \subseteq \mathbb{R}^n$.

We thus see that Bourgain's slicing problem can be formulated as a question on the relation between the covariance of a convex body and its volume. Note that the logarithm of the volume of a convex body is the differential entropy of a random vector X that is distributed uniformly over the convex body. In general, when the random vector X has density ρ in $\mathbb{R}^n$, its differential entropy is

$$Ent(X) = -\int_{\mathbb{R}^n} \rho \log \rho.$$

Definition 82. *For a convex body $K \subseteq \mathbb{R}^n$ we define its isotropic constant to be*

$$L_K = \left(\frac{\det \mathrm{Cov}(K)}{Vol_n(K)^2} \right)^{\frac{1}{2n}}$$

where $\mathrm{Cov}(K)$ is the covariance matrix of the uniform probability distribution on K. More generally, the isotropic constant of an absolutely continuous, log-concave random vector X in $\mathbb{R}^n$ is

$$L_X = \left(\frac{\det \mathrm{Cov}(X)}{e^{2Ent(X)}} \right)^{\frac{1}{2n}}. \tag{56}$$

The isotropic constant of a convex body $K \subseteq \mathbb{R}^n$ of volume one governs the volumes of its hyperplane sections. From Corollary 81 we see that when $Vol_n(K) = 1$, there always exists a hyperplane section $H \subseteq \mathbb{R}^n$ with

$$Vol_{n-1}(K \cap H) \geq c / L_K.$$

Moreover, if we additionally assume that $\mathrm{Cov}(K)$ is a scalar matrix, then for any hyperplane $H \subseteq \mathbb{R}^n$ through the barycenter of K,

$$Vol_{n-1}(K \cap H) \sim \frac{1}{L_K}.$$

The slicing problem thus asks whether L_K is universally bounded from above.

Remark on the definition of the isotropic constant in the log-concave case. Some variants of this definition exist, sometimes one replaces $Ent(X)$ by $-\log\sup\rho$ or by $-\log\rho(\mathbb{E}X)$ or by $2\log\mathbb{E}\rho^{-1/2}(X)$, where ρ is the density of X. See for instance [20] where this is discussed in more details. These variants differ at most by a multiplicative universal constant, because of the following lemma:

Lemma 83. *Denoting by $\psi = -\log\rho$ the convex potential of X, we have*

$$\psi(\mathbb{E}X) \leq Ent(X) \leq \inf\psi + n$$

and

$$\mathbb{E}e^{\frac{\psi(X)}{2}} \leq e^{\frac{\inf\psi}{2}+(\ln 2)n}.$$

Proof. We may assume that ρ is continuous in $\mathbb{R}^n$ in order to neglect boundary terms in the integration by parts below. Let $y \in \mathbb{R}^n$. Then by Jensen's inequality and by the fact that any convex function lies above its tangent at X,

$$\psi(\mathbb{E}X) \leq \mathbb{E}\psi(X) = Ent(X) = \mathbb{E}\psi(X) \leq \mathbb{E}\left[\psi(y) - \nabla\psi(X)\cdot(y-X)\right] = \psi(y) + n.$$

Additionally,

$$\mathbb{E}e^{\frac{\psi(X)}{2}} = e^{\frac{\psi(y)}{2}}\int_{\mathbb{R}^n} e^{-\frac{\psi(x)+\psi(y)}{2}}dx \leq e^{\frac{\psi(y)}{2}}\int_{\mathbb{R}^n} e^{-\psi\left(\frac{x+y}{2}\right)}dx = 2^n e^{\frac{\psi(y)}{2}}\int e^{-\psi} = 2^n e^{\frac{\psi(y)}{2}}.$$

The lemma follows by taking the infimum over all $y \in \mathbb{R}^n$ in these two inequalities. □

It what follows we work with the definition (56). While here we are interested only in the log-concave case, the definition makes sense for any absolutely continuous random vector X with finite second moments in $\mathbb{R}^n$. The isotropic constant measures the difference between two ways to measure the "size" of a random vector: its entropy and its covariance. Here are some basic properties of the isotropic constant:

1. It is an affine invariant, $L_{T(X)} = L_X$ for any invertible linear-affine map $T : \mathbb{R}^n \to \mathbb{R}^n$.

2. If $X_1, X_2 \in \mathbb{R}^n$ are independent log-concave random vectors, then for $X = (X_1, X_2) \in \mathbb{R}^{2n} \cong \mathbb{R}^n \times \mathbb{R}^n$,
$$L_X = \sqrt{L_{X_1}L_{X_2}}.$$

3. For any dimension n and an absolutely continuous random vector X with finite second moments in $\mathbb{R}^n$,
$$L_X \geq \frac{1}{\sqrt{2\pi e}},$$
with equality when X is Gaussian. Indeed, this amounts to showing that among all random vectors with a fixed covariance matrix in $\mathbb{R}^n$, the differential entropy is maximal for the Gaussian distribution, which is a standard fact. We recall the short proof below.

Proof. Suppose that X is a random vector of mean zero, density ρ, and let G be a centered Gaussian vector with the same covariance matrix as X. Also let γ be the density of G. Then

$$\mathbb{E}\log\left(\frac{\gamma(X)}{\rho(X)}\right) \leq \mathbb{E}\frac{\gamma(X)}{\rho(X)} - 1 = 0.$$

Since $\log\gamma$ is a quadratic function and X and G have the same covariance matrix, we get

$$Ent(X) = -\mathbb{E}\log\rho(X) \leq -\mathbb{E}\log\gamma(X) = -\mathbb{E}\log\gamma(G) = Ent(G). \qquad \square.$$

Exercise 9. Explain why it is not a coincidence that this universal constant $\sqrt{2\pi e}$ is "the same number" from the asymptotics $Vol_n(\sqrt{n}B^n)^{1/n} \approx 1/\sqrt{2\pi e}$.

4. Some examples:

$$L_{[0,1]^n} = \frac{1}{\sqrt{12}}, \qquad\qquad L_{\Delta^n} = \frac{(n!)^{1/n}}{(n+1)^{(n+1)/(2n)}\sqrt{n+2}} \approx \frac{1}{e}.$$

where Δ^n is a regular simplex in $\mathbb{R}^n$.

There are quite a few equivalent formulations and conditional statements, relating the isotropic constant to classical conjectures and results:

- If the isotropic constant is maximized for the cube among all centrally-symmetric convex set, then the Minkowski lattice conjecture follows, see Magazinov [66] and references therein. The Minkowski lattice conjecture suggests that if $L \subseteq \mathbb{R}^n$ is a lattice of determinant one, then each of its translates intersects the set

$$\left\{x \in \mathbb{R}^n\,;\, \prod_{i=1}^n |x_i| \leq \frac{1}{2^n}\right\}.$$

  This was proven in two dimensions by Minkowski in 1908.

- If the isotropic constant is maximized for the simplex among all convex bodies, then the Mahler conjecture follows in the non-symmetric case. This conjecture suggests that among all convex bodies $K \subseteq \mathbb{R}^n$, the volume product

$$Vol_n(K)\cdot Vol_n(K^\circ)$$

  is minimized when K is a centered simplex [52]. This was proven in two dimensions by Mahler in 1908. Here

$$K^\circ = \{x \in \mathbb{R}^n\,;\, \forall y \in K,\ x\cdot y \leq 1\}$$

  is the dual body. Recall that $(K^\circ)^\circ = K$ when K is a closed, convex set containing the origin. The Bourgain-Milman inequality resolves this conjecture up to a factor that is only exponential in the dimension. It states that for any convex body $K \subseteq \mathbb{R}^n$ containing the origin,

$$Vol_n(K)\cdot Vol_n(K^\circ) \geq (c/n)^n,$$

  for a universal constant $c > 0$.

- Suppose that $K \subseteq \mathbb{R}^n$ is a convex body. Is there an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ with $Vol_n(\mathcal{E}) = Vol_n(K)$ such that
$$Vol_n(K \cap C\mathcal{E}) \geq \frac{1}{2} \cdot Vol_n(K)$$
where $C > 0$ is a universal constant? This is an equivalent formulation of the slicing problem.

  Exercise 10. Prove the equivalence using reverse Hölder inequalities for quadratic polynomials.

  For any convex body $K \subseteq \mathbb{R}^n$, Milman's ellipsoid theorem [69] provides an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ with
$$Vol_n(K \cap C\mathcal{E}) \geq c^n \cdot Vol_n(K).$$
This suffices for developing the Milman ellipsoid theory, which contains the quotient of subspace theorem and reverse Brunn-Minkowski and the Bourgain-Milman inequality. See Pisier [74] and references therein. The slicing problem is a conjectural strengthening of Milman's ellipsoids.

We move on to discuss the $\sqrt{\log}$-bound for the isotropic constant, and the relation to the Poincaré constant and the thin shell constants. We define

$$\sigma_n = \sup_X \sqrt{\mathrm{Var}(|X|^2)/n}$$

where the supremum ranges over all isotropic, log-concave random vectors X in $\mathbb{R}^n$. By reverse Hölder inequalities for polynomials we may show that $\mathrm{Var}(|X|^2)/n \sim \mathrm{Var}(|X|)$, and hence σ_n is roughly the maximal width of the thin spherical shell that captures most of the mass of an isotropic, log-concave random vector.

From Corollary 42 we know that,

$$\sigma_n \leq \sup_X \sqrt{C_P(X) \cdot 4\mathbb{E}|X|^2/n} \leq \sup_X 2\sqrt{C_P(X)} \leq C\sqrt{\log n}.$$

Hence it remains to prove:

Theorem 84 (Eldan, Klartag [36])**.** *For any convex body $K \subseteq \mathbb{R}^n$,*

$$L_K \leq C\sigma_n.$$

Remark 85. In fact, it is shown in [36] that $L_X \leq C\sigma_n$ for any log-concave random vector X in $\mathbb{R}^n$, but for simplicity we confine ourselves here for the convex body case. The slicing problem for convex bodies and for log-concave measures are known to be equivalent, as shown by Ball [7, 47].

While we studied Gaussian convolution in sections 5 and 6, the proof of Theorem 84 utilizes the closely related *Laplace transform*. Let us fix an isotropic, log-concave random vector X with density ρ in $\mathbb{R}^n$. Its logarithmic Laplace transform is

$$\Lambda(y) = \Lambda_X(y) = \log \mathbb{E} e^{X \cdot y}.$$

Since a log-concave random vector has exponential moments, the logarithmic Laplace transform is finite near the origin. In fact, it is smooth in the open convex set $\Omega = \{\Lambda < \infty\}$. For $y \in \Omega$ we write X_y for a random vector with density

$$\rho_y(x) = \frac{\rho(x) e^{x \cdot y}}{e^{\Lambda(y)}}.$$

It is again a log-concave random vector, not necessarily isotropic, and we think of it as a *tilted* version of the random vector X. We comment that it is possible to view tilts using projective transformations, this leads to the conditional statement that the strong slicing conjecture implies the Mahler conjecture, see [52].

Lemma 86. *For any $y \in \Omega$,*

$$\nabla \Lambda(y) = \mathbb{E} X_y, \qquad \nabla^2 \Lambda(y) = \mathrm{Cov}(X_y), \qquad \nabla^3 \Lambda(y) = \mathbb{E}(X_y - a_y)^{\otimes 3},$$

where $a_y = \mathbb{E} X_y$.

Lemma 86 is proven by direct computation; the logarithmic Laplace transform is the cumulant generating function. We see from Lemma 86 that Λ is convex, even strongly-convex as its Hessian is positive definite. In particular the gradient $\nabla \Lambda : \Omega \to \mathbb{R}^n$ is a one-to-one map. Consider the "tilted determinant" function

$$F(y) = \log \det \nabla^2 \Lambda(y) = \log \det \mathrm{Cov}(X_y).$$

It measures how the determinant of the covariance matrix changes when we tilt the given distribution. Occasionally we may view F as a function that is defined only up to an additive constant. Write $[F]$ for the equivalence class of F under the equivalence relation "F is equivalent to G if and only if $F - G$ is a constant function".

Lemma 87. *The following bound holds pointwise in all of Ω:*

$$(\nabla^2 \Lambda)^{-1} \nabla F \cdot \nabla F \leq n \sigma_n^2. \tag{57}$$

Proof. Let us prove this bound first for $y = 0$ using the isotropicity of X. Recalling how to differentiate a determinant, we see that for any unit vector $v \in S^{n-1}$,

$$\partial_v F(0) = \mathrm{Tr}\left[(\nabla^2 \Lambda)^{-1}(0) \cdot \partial_v \nabla^2 \Lambda(0) \right] = \mathbb{E}(X \cdot v)|X|^2 \leq \sqrt{\mathbb{E}(X \cdot v)^2 \cdot \mathrm{Var}(|X|^2)} \leq \sqrt{n} \sigma_n.$$

By considering the supremum over all $v \in S^{n-1}$, we obtain the desired bound at $y = 0$.

In order to obtain the bound for any $y \in \Omega$ we may either make a computation, or alternatively, think invariantly without computing anything, as we now explain.

Define a Riemannian metric on Ω via the Hessian of the log-Laplace transform Λ. We look at the Hessian metric (Ω, g), where the scalar product of two tangent vectors $u, v \in T_x\mathbb{R}^n \cong \mathbb{R}^n$ is

$$g_x(u, v) = \nabla^2\Lambda(x)u \cdot v.$$

The main observation is that the expression on the left-hand side of (57) is the squared Riemannian length of the Riemannian gradient of the function $F : \Omega \to \mathbb{R}$. We say that

$$\mathcal{M}_X = (\Omega, g, [F])$$

is the "Riemannian package" associated with X. This means that (Ω, g) is a Riemannian manifold and that F is a function on Ω modulo an additive constant. An isomorphism between two Riemannian packages is a bijective map which is a Riemannian isometry and transforms correctly the function modulo the additive constant.

What happens to the Riemannian package associated with X when we do various operations?

- When we translate X, the Riemannian metric stays the same, as well as the function F. We get the same Riemannian package.
- Tilting X and switching to X_y yields an isomorphism of the two Riemannian packages by *translation* by y: We translate Ω, g and $[F]$ by the vector $y \in \Omega$. Any translation corresponds to a tilt and vice versa.
- Applying an invertible linear transformation to X induces an isomorphism of the Riemannian packages. We apply a linear transformation and push forward Ω, g and $[F]$. (See also the paragraph before the next lemma).

By the first and last items, we proved (57) at the point $y = 0$ for any log-concave random vector (not necessarily centered or isotropic). By the middle item, we proved (57) also at all other points of Ω. We refer to [36] for a more detailed proof. □

It makes sense to say that we think of X as a random vector defined on an abstract affine space, rather than on $\mathbb{R}^n$, and observe that the Riemannian manifold (Ω, g) is well-defined, as well as the function $F : \Omega \to \mathbb{R}$ modulo additive constants. What can we say about balls in this Riemannian manifold?

Lemma 88. *Assume that X is a centered, log-concave random vector in $\mathbb{R}^n$. Then for any $r > 0$,*

$$\frac{1}{2} \cdot \{\Lambda \leq r\} \subseteq B_g(0, \sqrt{r}).$$

Proof. Let $y \in \Omega$ satisfy $\Lambda(2y) \leq r$. We need to find a curve from 0 to y whose Riemannian length is at most r. Let us try a line segment:

$$\begin{aligned} Length_g([0,y]) &= \int_0^1 \sqrt{\nabla^2\Lambda(ty)y \cdot y}dt = \int_0^1 \sqrt{\frac{d^2}{dt^2}\Lambda(ty)}dt \\ &\leq \sqrt{\int_0^2 (2-t)\frac{d^2}{dt^2}\Lambda(ty)dt \cdot \int_0^1 \frac{1}{2-t}dt} \\ &= \sqrt{\log 2} \cdot \sqrt{\Lambda(2y) - [\Lambda(0) + \nabla\Lambda(0)\cdot(2y)]} \\ &= \sqrt{\log 2} \cdot \sqrt{\Lambda(2y)} \leq \sqrt{r}. \end{aligned}$$

□

Let X be an isotropic random vector in $\mathbb{R}^n$, distributed uniformly in a convex body $K \subseteq \mathbb{R}^n$. We need two estimates for the proof of Theorem 84:

(i) First, we need to show that for $r = n/\sigma_n^2$,

$$Vol_n(K) \geq e^{-n} \cdot Vol_n(B_g(0,\sqrt{r})),$$

the Euclidean volume of the Riemannian ball. This is related to mass transport in a simple case.

(ii) Second, we need to show that

$$Vol_n(\{\Lambda \leq r\})^{1/n} \geq c\frac{r}{n}L_K.$$

This is related to the Bourgain-Milman inequality.

Proof of Theorem 84. Since X is isotropic and log-concave, by (i), (ii) and Lemma 88,

$$\begin{aligned} L_K = Vol_n(K)^{-1/n} &\leq C \cdot Vol_n(B_g(0,\sqrt{r}))^{-1/n} \\ &\leq 2C \cdot Vol_n(\{\Lambda \leq r\})^{-1/n} \leq C'\frac{n}{rL_K} = C'\frac{\sigma_n^2}{L_K}. \end{aligned}$$

Thus $L_K \leq C'' \cdot \sigma_n$. □

Proof of estimate (i): The function F vanishes at the origin, and by Lemma 87 it is a Riemannian Lipschitz function with Lipschitz constant at most $\sqrt{n}\sigma_n$. Hence,

$$|F| \leq n \qquad \text{in} \qquad B_g(0,\sqrt{r}).$$

Consequently, for any $y \in B_g(0,\sqrt{r})$,

$$e^{-n} \leq \det \nabla^2\Lambda(y) \leq e^n.$$

We will use the fact that $\nabla\Lambda(y) = \mathbb{E}X_y \in K$ and that $y \mapsto \nabla\Lambda(y)$ is one-to-one. Changing variables, we obtain

$$Vol_n(K) \geq Vol_n\left(\nabla\Lambda(B_g(0,\sqrt{r}))\right) = \int_{B_g(0,\sqrt{r})} \det \nabla^2\Lambda(y)\,dy \geq e^{-n} \cdot Vol_n(B_g(0,\sqrt{r})).$$ □

Proof of estimate (ii): For any $y \in rK^\circ$,

$$\Lambda(y) = \log \mathbb{E}e^{y \cdot X} \leq \log(e^r) = r.$$

Therefore,

$$\{\Lambda \leq r\} \supseteq rK^\circ.$$

By the Bourgain-Milman inequality,

$$Vol_n(\{\Lambda \leq r\})^{1/n} \geq Vol_n(rK^\circ)^{1/n} \geq c\frac{r}{n} Vol_n(K)^{-1/n} = c\frac{r}{n} L_K. \qquad \square$$

We remark that the Bourgain-Milman inequality has several proofs, and in particular it may be proven using more delicate analysis of the log-Laplace transform as shown by Giannopoulos, Paouris and Vritsiou [40].

References

[1] R. Adamczak, R. Latała, A. E. Litvak, K. Oleszkiewicz, A. Pajor, and N. Tomczak-Jaegermann. A short proof of Paouris' inequality. *Can. Math. Bull.*, 57(1):3–8, 2014.

[2] G. Allaire. À la recherche de l'inégalité perdue. *Matapli*, 98:52–64, 2012.

[3] L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Math.*, pages 1–52. Springer, Berlin, 2003.

[4] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman. *Asymptotic geometric analysis. Part II*, volume 261 of *Math. Surv. Monogr.* Providence, RI: American Mathematical Society (AMS), 2021.

[5] D. Bakry and M. Emery. Hypercontractivité de semi-groupes de diffusion (Hypercontractivity for diffusion semi-groups). *C. R. Acad. Sci., Paris, Sér. I*, 299:775–778, 1984.

[6] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren Math. Wiss.* Cham: Springer, 2014.

[7] K. Ball. Logarithmically concave functions and sections of convex sets in $\mathbf{R}^n$. *Studia Math.*, 88(1):69–84, 1988.

[8] K. Ball. The legacy of Jean Bourgain in geometric functional analysis. *Bull. Am. Math. Soc., New Ser.*, 58(2):205–223, 2021.

[9] K. Ball and V. H. Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math.*, 213(1):81–96, 2012.

[10] R. Bauerschmidt, T. Bodineau, and B. Dagallier. Stochastic dynamics and the Polchinski equation: an introduction. *Probab. Surv.*, 21:200–290, 2024.

[11] V. Bayle. *Propriétés de concavité du profil isopérimétrique et applications*. PhD thesis, Université Joseph Fourier Grenoble, 2003.

[12] P. Bizeul. On the log-Sobolev constant of log-concave measures. Preprint, arXiv:2306.12997 [math.FA] (2023), 2023.

[13] P. Bizeul. On measures strongly log-concave on a subspace. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 60(2):1090–1100, 2024.

[14] S. G. Bobkov. Extremal properties of half-spaces for log-concave distributions. *Ann. Probab.*, 24(1):35–48, 1996.

[15] S. G. Bobkov. Remarks on the growth of L^p-norms of polynomials. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 27–35. Springer, Berlin, 2000.

[16] S. G. Bobkov. On isoperimetric constants for log-concave probability distributions. In *Geometric aspects of functional analysis. Proceedings of the Israel seminar (GAFA) 2004-2005*, pages 81–88. Berlin: Springer, 2007.

[17] S. G. Bobkov, G. Chistyakov, and F. Götze. *Concentration and Gaussian approximation for randomized sums*, volume 104 of *Probability Theory and Stochastic Modelling*. Springer, 2023.

[18] S. G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 80(7):669–696, 2001.

[19] S. G. Bobkov and C. Houdré. Isoperimetric constants for product probability measures. *Ann. Probab.*, 25(1):184–205, 1997.

[20] S. G. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inf. Theory*, 57(8):4940–4954, 2011.

[21] V. I. Bogachev and A. V. Kolesnikov. The Monge-Kantorovich problem: achievements, connections, and perspectives. *Russ. Math. Surv.*, 67(5):785–890, 2012.

[22] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30:207–216, 1975.

[23] J. Bourgain. On high-dimensional maximal functions associated to convex bodies. *Amer. J. Math.*, 108(6):1467–1476, 1986.

[24] H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Functional Analysis*, 22(4):366–389, 1976.

[25] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of isotropic convex bodies*, volume 196 of *Math. Surv. Monogr.* Providence, RI: American Mathematical Society (AMS), 2014.

[26] P. Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)*, 15(2):213–230, 1982.

[27] L. A. Caffarelli, M. Feldman, and R. J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.*, 15(1):1–26, 2002.

[28] F. Cavalletti and A. Mondino. Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds. *Invent. Math.*, 208(3):803–849, 2017.

[29] A. Cayley. On Monge's mémoire sur la théorie des déblais et des remblais. *Proc. Lond. Math. Soc.*, 14:139–142, 1883.

[30] D. Chafaï and J. Lehec. On Poincaré and logarithmic Sobolev inequalities for a class of singular Gibbs measures. In *Geometric aspects of functional analysis. Israel seminar (GAFA) 2017–2019. Volume 1*, pages 219–246. Cham: Springer, 2020.

[31] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in analysis, Princeton Univ. Press*, pages 195–199, 1970.

[32] Y. Chen. An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geom. Funct. Anal.*, 31(1):34–61, 2021.

[33] N. De Ponti and A. Mondino. Sharp Cheeger-Buser type inequalities in RCD(K, ∞) spaces. *J. Geom. Anal.*, 31(3):2416–2438, 2021.

[34] P. Diaconis and D. Freedman. Asymptotics of graphical projection pursuit. *Ann. Statist.*, 12(3):793–815, 1984.

[35] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom. Funct. Anal.*, 23(2):532–569, 2013.

[36] R. Eldan and B. Klartag. Approximately Gaussian marginals and the hyperplane conjecture. In *Concentration, functional inequalities and isoperimetry*, volume 545 of *Contemp. Math.*, pages 55–68. Amer. Math. Soc., Providence, RI, 2011.

[37] L. C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137(653):viii+66, 1999.

[38] M. Fradelizi. Hyperplane sections of convex bodies in isotropic position. *Beiträge Algebra Geom.*, 40(1):163–183, 1999.

[39] D. A. Freedman. On tail probabilities for martingales. *Ann. Probab.*, 3:100–118, 1975.

[40] A. Giannopoulos, G. Paouris, and B.-H. Vritsiou. The isotropic position and the reverse Santaló inequality. *Israel J. Math.*, 203(1):1–22, 2014.

[41] N. Gozlan, C. Roberto, and P.-M. Samson. From concentration to logarithmic Sobolev and Poincaré inequalities. *J. Funct. Anal.*, 260(5):1491–1522, 2011.

[42] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *American Journal of Mathematics*, 105(4):843–854, 1983.

[43] O. Guédon and E. Milman. Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geom. Funct. Anal.*, 21(5):1043–1068, 2011.

[44] D. Hensley. Slicing convex bodies—bounds for slice area in terms of the body's covariance. *Proc. Amer. Math. Soc.*, 79(4):619–625, 1980.

[45] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

[46] L. V. Kantorovich and G. P. Akilov. *Functional analysis*. Pergamon Press, Oxford-Elmsford, N.Y., second edition, 1982. Translated from the Russian by Howard L. Silcock.

[47] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. Funct. Anal.*, 16(6):1274–1290, 2006.

[48] B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131, 2007.

[49] B. Klartag. A Berry-Esseen type inequality for convex bodies with an unconditional basis. *Probab. Theory Related Fields*, 145(1-2):1–33, 2009.

[50] B. Klartag. Needle decompositions in Riemannian geometry. *Mem. Amer. Math. Soc.*, 249(1180):v+77, 2017.

[51] B. Klartag. Eldan's stochastic localization and tubular neighborhoods of complex-analytic sets. *J. Geom. Anal.*, 28(3):2008–2027, 2018.

[52] B. Klartag. Isotropic constants and Mahler volumes. *Adv. Math.*, 330:74–108, 2018.

[53] B. Klartag. On Yuansi Chen's work on the KLS conjecture. *Lecture notes prepared for a winter school at the Hausdorff Institute*, 2021. Available at https://www.him.uni-bonn.de/fileadmin/him/Lecture_Notes/chen_lecture_not

[54] B. Klartag. Logarithmic bounds for isoperimetry and slices of convex sets. *Ars Inveniendi Analytica*, Paper No. 4, 17pp, 2023.

[55] B. Klartag and J. Lehec. Bourgain's slicing problem and KLS isoperimetry up to polylog. *Geom. Funct. Anal.*, 32(5):1134–1159, 2022.

[56] B. Klartag and O. Ordentlich. The strong data processing inequality under the heat flow. Preprint, arXiv:2406.03427 [cs.IT] (2024), 2024.

[57] B. Klartag and E. Putterman. Spectral monotonicity under Gaussian convolution. *Ann. Fac. Sci. Toulouse, Math. (6)*, 32(5):939–967, 2023.

[58] M. Ledoux. A simple analytic proof of an inequality by P. Buser. *Proc. Am. Math. Soc.*, 121(3):951–959, 1994.

[59] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[60] M. Ledoux. From concentration to isoperimetry: semigroup proofs. In *Concentration, functional inequalities and isoperimetry. Mainly the proceedings of the international workshop on concentration, functional inequalities and isoperimetry, Florida Atlantic University, Boca Raton, FL, USA, October 29–November 1, 2009.*, pages 155–166. Providence, RI: American Mathematical Society (AMS), 2011.

[61] M. Ledoux. From concentration to isoperimetry: semigroup proofs. In *Concentration, functional inequalities and isoperimetry. functional inequalities and isoperimetry, Florida Atlantic University, Boca Raton, FL, USA, October 29–November 1, 2009.*, pages 155–166. Providence, RI: American Mathematical Society (AMS), 2011.

[62] Y. T. Lee and S. S. Vempala. The Kannan-Lovász-Simonovits conjecture. In *Current developments in mathematics 2017*, pages 1–36. Int. Press, Somerville, MA, 2019.

[63] Y. T. Lee and S. S. Vempala. Eldan's stochastic localization and the KLS conjecture: Isoperimetry, concentration and mixing. *Ann. of Math. (2)*, 199(3):1043–1092, 2024.

[64] A. Lichnerowicz. *Géométrie des groupes de transformations*, volume III of *Travaux et Recherches Mathématiques*. Dunod, Paris, 1958.

[65] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14:860–872, 1986.

[66] A. Magazinov. A proof of a conjecture by Haviv, Lyubashevsky and Regev on the second moment of a lattice Voronoi cell. *Adv. Geom.*, 20(1):117–120, 2020.

[67] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1):1–43, 2009.

[68] E. Milman. Isoperimetric and concentration inequalities: equivalence under curvature lower bound. *Duke Math. J.*, 154(2):207–239, 2010.

[69] V. D. Milman. An inverse form of the Brunn-Minkowski inequality with applications to local theory of normed spaces. *C. R. Acad. Sci., Paris, Sér. I*, 302:25–28, 1986.

[70] S.-I. Ohta. Needle decompositions and isoperimetric inequalities in Finsler geometry. *J. Math. Soc. Japan*, 70(2):651–693, 2018.

[71] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[72] G. Paouris. Concentration of mass on convex bodies. *Geom. Funct. Anal.*, 16(5):1021–1049, 2006.

[73] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292, 1960.

[74] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[75] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.

[76] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2: Itô calculus.* Cambridge: Cambridge University Press, 2nd ed. edition, 2000.

[77] O. S. Rothaus. Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 64:296–313, 1985.

[78] V. N. Sudakov. Typical distributions of linear functionals in finite-dimensional spaces of high dimension. *Dokl. Akad. Nauk SSSR*, 243(6):1402–1405, 1978. English translation: Soviet Math. Dokl. 19 (1978), no. 6, 1578–1582 (1979).

[79] V. N. Sudakov and B. S. Cirel'son. Extremal properties of half-spaces for spherically invariant measures. *J. Sov. Math.*, 9:9–18, 1978.

[80] G. Szegö. Inequalities for certain eigenvalues of a membrane of given area. *J. Rational Mech. Anal.*, 3:343–356, 1954.

[81] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

[82] N. Th. Varopoulos. Small time Gaussian estimates of heat diffusion kernels. I: The semigroup technique. *Bull. Sci. Math., II. Sér.*, 113(3):253–277, 1989.

[83] H.F. Weinberger. An isoperimetric inequality for the n-dimensional free membrane problem. *J. Rational Mech. Anal.*, 5:633–636, 1956.