Lecture notes MSRI Summer school on Concentration and Localization methods in Probability and Geometry

Max Fathi *   Marcel Hudiani †   Jiayi Wang ‡

August 14, 2023

1 Introduction

This document contains the notes for a summer school course given by the first named author during an MSRI/SLMath summer school entitled "Concentration Inequalities and Localization Techniques in High Dimensional Probability and Geometry", that took place in July 2023.

It was one of two courses, the other one was given by Dan Mikulincer. The two courses were intertwined, and several notions that are used in these notes were presented in the other course, including Poincaré inequalities and the Kannan-Lovesz-Simonovits bisection method for building needle decompositions on convex domains of Euclidean spaces. Some aspects were more developed during the lectures (notably towards the end of these notes, in particular solving the Monge problem for the distance cost, and the construction of needle decompositions). Additionally, the lectures were completed by exercise sessions that were run by Arianna Piana and Shay Sadovsky.

2 Brunn-Minkowski inequality and applications

The content of this section is based on the first chapter of [2]. We refer to it for more about the Brunn-Minkowski, including other proofs and applications, as well as the history of the problem.

2.1 Brunn-Minkowski inequality

Theorem 1 (Brunn-Minkowski inequality)**.** *Let $A, B \subset \mathbb{R}^d$ be two compact non-empty sets. Then*

$$\mathrm{Vol}_d(A+B)^{1/d} \geq \mathrm{Vol}_d(A)^{1/d} + \mathrm{Vol}_d(B)^{1/d}.$$

This inequality can also be viewed as a type of concavity property, when rewritten as

$$\mathrm{Vol}_d(\lambda A + (1-\lambda)B)^{1/d} \geq \lambda \,\mathrm{Vol}_d(A)^{1/d} + (1-\lambda)\,\mathrm{Vol}_d(B)^{1/d}$$

for any $\lambda \in [0,1]$.

*LJLL and LPSM, Université Paris Cité, mfathi@lpsm.paris
†Department of Mathematics, University of Arizona, marcelh@math.arizona.edu
‡Department of Mathematics, Duke University, jw771@duke.edu

We will mostly be interested, and only prove, the case where A and B are convex bodies, that is compact, convex sets with non-empty interior.

An equivalent formulation is the multiplicative form

$$\mathrm{Vol}_d(\lambda A + (1-\lambda)B) \geq \mathrm{Vol}_d(A)^{\lambda} \times \mathrm{Vol}_d(B)^{1-\lambda} \tag{1}$$

which has the advantage (or,depending on the context, inconvenience) of not explicitly depending on the dimension. It can be deduced from the previous statement by applying the arithmetic-geometric inequality. However, it is actually equivalent: assuming without loss of generality that both sets have (strictly) positive volume, consider

$$A_1 := \mathrm{Vol}_d(A)^{-1/d}A; \quad B_1 := \mathrm{Vol}_d(B)^{-1/d}B; \quad \lambda := \frac{\mathrm{Vol}_d(A)^{1/d}}{\mathrm{Vol}_d(A)^{1/d} + \mathrm{Vol}_d(B)^{1/d}}.$$

Both A_1 and B_1 have volume 1, and hence if we apply to them the multiplicative version,

$$\mathrm{Vol}_d(\lambda A_1 + (1-\lambda)B_1) \geq 1.$$

But since

$$\lambda A_1 + (1-\lambda)B_1 = \frac{A+B}{\mathrm{Vol}_d(A)^{1/d} + \mathrm{Vol}_d(B)^{1/d}}$$

we immediately get the arithmetic form.

Note also that with this argument, we see it is enough to prove the Brunn-Minkowski inequality for sets with volume 1, which will be useful in the sequel.

Proof of the Brunn-Minkowski inequality for convex sets. We shall proceed by induction on the dimension. The case $d = 1$ is immediate, since

$$\lambda[a, b] + (1-\lambda)[c, d] = [\lambda a + (1-\lambda)c; \lambda b + (1-\lambda)d].$$

Let us assume the statement is true for all dimensions less than $d - 1$. Let K_0 and K_1 be two convex bodies in dimension d. Without loss of generality, we can assume their volume to be equal to 1.

Let $\theta \in \mathbb{S}^{d-1}$. We define for $i \in \{1, 2\}$

$$f_i(t) := \mathrm{Vol}_{d-1}(\{x \in K_i; \langle x, \theta\rangle = t\})$$
$$g_i(t) := \mathrm{Vol}_{d-1}(\{x \in K_i; \langle x, \theta\rangle \leq t\})$$

If X is a random variable uniformly distributed on K_i, thet f_i is the density of the variable $\langle X, \theta\rangle$ while g_i is its cumulative distribution function. Let $[a_i, b_i]$ be the support of f_i. We also define $h_i : (0, 1) \longrightarrow \mathbb{R}$ to be the inverse function of g_i. It is differentiable, and

$$h_i'(u) = \frac{1}{g'(h_i(u))} = \frac{1}{f_i(h_i(u))}.$$

Let $h_\lambda = (1-\lambda)h_0 + \lambda h_1$ and $K_\lambda = (1-\lambda)K_0 + \lambda K_1$. Let

$$K_\lambda(u) := K\lambda \cap \{y + h_\lambda(u)\theta; y \perp \theta\}$$

be the (reparametrized) decomposition of K_λ into slices along the direction θ. As an immediate consequence of these definitions,

$$(1-\lambda)K_0(u) + \lambda K_1(u) \subset K_\lambda(u).$$

By making the change of variable $t = h_\lambda(u)$ on the range of h_λ, we have

$$\begin{aligned}
\mathrm{Vol}_d(K_\lambda) &= \int \mathrm{Vol}_{d-1}(K_\lambda \cap \{t\theta + y; y \perp \theta\})dt \\
&\geq \int_0^1 \mathrm{Vol}_{d-1}(K_\lambda(u))h'_\lambda(u)du \\
&= \int_0^1 \mathrm{Vol}_{d-1}(K_\lambda(u)) \left(\frac{1-\lambda}{f_0(h_0(u))} + \frac{\lambda}{f_1(h_1(u))} \right) du \\
&\geq \int_0^1 \mathrm{Vol}_{d-1}((1-\lambda)K_0(u) + \lambda K_1(u)) \left(\frac{1-\lambda}{f_0(h_0(u))} + \frac{\lambda}{f_1(h_1(u))} \right) du \\
&\geq \int_0^1 f_0(h_0(u))^{1-\lambda} f_1(h_1(u))^\lambda \left(\frac{1-\lambda}{f_0(h_0(u))} + \frac{\lambda}{f_1(h_1(u))} \right) du.
\end{aligned}$$

Using the arithmetic-geometric inequality on the second factor we see that the integrand is bounded from below by 1, which concludes the proof. □

Let's look at the equality cases in this proof. Due to the translation invariance, we can assume that both K_0 and K_1 have their barycenter at the origin. Equality in the Brunn-Minkowski inequality requires equality at the point where we used the arithmetic-geometric inequality. Hence if there is equality, then $f_0 \circ h_0 = f_1 \circ h_1$, and hence $h'_0 = h'_1$. Therefore $h_1 - h_0$ is constant. But since the barycenter is at the origin, then

$$0 = \int_{K_i} \langle x, \theta \rangle dx = \int t f_i(t) dt$$

$$= \int_0^1 h_i(u) f_i(h_i(u)) h'_i(u) du = \int h_i(u) du$$

so that actually $h_1 = h_0$. Therefore $g_0 = g_1$ and the boundaries of the two convex sets must be the same, since the extremal in every direction θ is the same.

For the general Brunn-Minkowski inequality (that allows for non-convex sets), equality holds iff the two sets are homothetic to subsets of full measures of a same convex set. See [16] for a stable version of this statement.

2.2 Applications

Theorem 2 (Borell's Lemma)**.** *Let K be a convex body with volume 1, and A be a symmetric convex closed subset of K, such that $\mathrm{Vol}_d(K \cap A) = \delta > 1/2$. Then for any $t > 1$ we have*

$$\mathrm{Vol}_d(K \cap (tA)^c) \leq \delta \left(\frac{1-\delta}{\delta} \right)^{(t+1)/2}.$$

Remark 3. *The constant 1/2 for the minimal value of the volume is not so important, we still get exponential decay in t for other values, up to changing other constants.*

Proof. We first show by contradiction that

$$\frac{2}{t+1}(tA)^c + \frac{t-1}{t+1}A \subset A^c.$$

If this is not so, there is $y \notin tA$ and $a, b \in A$ such that

$$a = \frac{2}{t+1} y + \frac{t-1}{t+1} b.$$

But then

$$\frac{y}{t} = \frac{t+1}{2t} a + \frac{t-1}{2t} (-b) \in A$$

by convexity and symmetry of A, which is a contradiction.

As a consequence,

$$\frac{2}{t+1} ((tA)^c \cap K) + \frac{t-1}{t+1} (A \cap K) \subset A^c \cap K.$$

Applying the Brunn-Minkowski inequality then yields

$$\begin{aligned} 1 - \delta = \mathrm{Vol}_d(A^c \cap K) &\geq \mathrm{Vol}_d((tA)^c \cap K)^{2/(t+1)} \mathrm{Vol}_d(A \cap K)^{(t-1)/(t+1)} \\ &= \mathrm{Vol}_d((tA)^c \cap K)^{2/(t+1)} \delta^{(t-1)/(t+1)}. \end{aligned}$$

Rearranging the terms concludes the proof. □

Definition 4. *The perimeter (Minkowski content) of a convex set K can be defined as*

$$\mathrm{Per}(K) := \liminf_{t \to 0^+} \frac{\mathrm{Vol}_d(K + tB_2^d) - \mathrm{Vol}_d(K)}{t}.$$

There are other common definitions, which coincide with this one.

Theorem 5 (The isoperimetric inequality)**.** *Let K be a convex body in $\mathbb{R}^d$. Then*

$$\mathrm{Per}(K) \geq d \, \mathrm{Vol}_d(B_2^d)^{1/d} \, \mathrm{Vol}_d(K)^{(d-1)/d}.$$

The isoperimetric inequality is true for more general sets, but one must pay attention to the definition of the boundary.

Proof. Take r such that $\mathrm{Vol}_d(K) = \mathrm{Vol}_d(rB_2^d)$. Then

$$\mathrm{Vol}_d(K + tB_2^d)^{1/d} \geq \mathrm{Vol}_d(K)^{1/d} + t \, \mathrm{Vol}_d(B_2^d)^{1/d} = (r+t) \, \mathrm{Vol}_d(B_2^d)^{1/d}$$

so that

$$\begin{aligned} \mathrm{Per}(K) &\geq \liminf \frac{(r+t)^d - t^d}{t} \mathrm{Vol}_d(B_2^d) \\ &= dr^{d-1} \, \mathrm{Vol}_d(B_2^d) \end{aligned}$$

and the conclusion follows since $r = (\mathrm{Vol}_d(K) / \mathrm{Vol}_d(B))^{1/d}$. □

3 Lecture 2: Prékopa-Leindler inequality and Gaussian logarithmic Sobolev inequality

3.1 Log-concave measures

Definition 6. *A non-negative measure μ on $\mathbb{R}^d$ is said to be log-concave if it satisfies the multiplicative form of the Brunn-Minkowski inequality, that is for any compact non-empty sets A and B we have*

$$\mu((1-\lambda)A + \lambda B) \geq \mu(A)^{1-\lambda} \mu(B)^{\lambda}.$$

Remark 7. *In general, such measures are* not *translation-invariant, and do not satisfy a scaling property with respect to dilations. In particular, the multiplicative form does* not *imply the additive form of the Brunn-Minkowski inequality for general measures. Optimal additive Brunn-Minkowski inequalities for more general measures are not so well-understood at this time. For example, the optimal exponent $1/d$ for the Gaussian additive Brunn-Minkowski inequality has only been established in 2021 by Eskenazis and Moschidis, answering a question of Gardner and Zvavitch, and only holds when restricting to convex, symmetric sets (Nayar and Tkocz gave a counterexample when symmetry is omitted).*

Since for proving Borell's lemma we only used the multiplicative form, it can be extended to log-concave measures, and we get

Theorem 8. *Let μ be a log-concave probability measure on $\mathbb{R}^d$. Then for any symmetric convex set A with $\mu(A) = \delta \in (1/2, 1)$ and any $t > 1$, we have*

$$1 - \mu(tA) \leq \delta \left(\frac{1-\delta}{\delta} \right)^{(t+1)/2}.$$

However, the isoperimetric inequality uses the additive form (it is dimensional), and hence does not generalize.

3.2 Prékopa-Leindler inequality

A common theme in geometric functional analysis is that many inequalities comparing volumes of sets have functional forms. In the case of the Brunn-Minkowski inequality, the functional form is the Prékopa-Leindler inequality:

Theorem 9 (Prékopa-Leindler inequality)**.** *Let f, g and h be three functions on $\mathbb{R}^d$, and $t \in (0, 1)$, such that*

$$h(tx + (1-t)y) \geq tf(x) + (1-t)g(y)$$

for all x, y in $\mathbb{R}^d$. Then

$$\int e^h dx \geq \left(\int e^f dx \right)^t \left(\int e^g dx \right)^{1-t}.$$

Corollary 10 (Measures with log-concave densities are log-concave)**.** *Let μ be a positive measure on $\mathbb{R}^d$, admitting a log-concave density w.r.t. the Lebesgue measure. Then it is log-concave, in the sense of Definition 6.*

Proof. Let e^V be the density, with V concave, and let A and B be two closed convex sets. Taking $f = V$ on A and $-\infty$ outside, g the same for B, and h the same on $tA + (1-t)B$, we immediately get the desired result by applying the Prékopa-Leindler inequality. □

The converse is also true, in a strong sense (see [6] for a proof):

Theorem 11 (Borell's theorem)**.** *If a log-concave probability measure on $\mathbb{R}^d$ is not supported on a hyperplane (that is, $\mu(H) < 1$ for any hyperplane H), then it is absolutely continuous w.r.t. the Lebesgue measure, and its density is log-concave.*

Corollary 12 (Convolutions of log-concave functions are log-concave)**.** *Let V and W be convex functions. Then*

$$z \longrightarrow -\log \int e^{-V(x)-W(z-x)} dx$$

is convex.

Proof. Apply the Prékopa-Leindler inequality to $f(x) = -V(x) - W(z_1 - x)$, $g(x) = -V(x) - W(z_2 - x)$ and $h(x) = -V(x) - W(tz_1 + (1-t)z_2 - x)$. □

Corollary 13 (Heat flow preserves log-concavity)**.** *Let ρ_0 be log-concave and L^1, and let $\rho(t,x)$ be the solution to*

$$\partial_t \rho = \Delta \rho.$$

Then for any $t > 0$, $\rho(t, \cdot)$ is log-concave.

Proof. The heat flow can be obtained by convolution with a time-dependent Gaussian kernel, with covariance matrix $2t\,\mathrm{Id}$. We can then apply the previous corollary. □

Proof of the Prékopa-Leindler inequality. We shall once again proceed by induction on the dimension. Let's start with dimension one. Without loss of generality, assume that f and g are continuous, positive probability densities. The goal is then to prove that $\int h \geq 1$.

We can reparametrize by the cumulative distribution functions, that is let $x(t)$, $y(t)$ be functions on $(0,1)$ defined by

$$\int_{-\infty}^{x(t)} f = t; \qquad \int_{-\infty}^{y(t)} g = t.$$

We have

$$x'(t) f(x(t)) = y'(t) g(y(t)) = 1.$$

Let

$$z(t) := \lambda x(t) + (1-\lambda) y(t),$$

which is a strictly increasing, C^1 function satisfying

$$\begin{aligned} z'(t) &= \lambda x'(t) + (1-\lambda) y'(t) \\ &\geq (x'(t))^{\lambda} (y'(t))^{1-\lambda)} \\ &= f(x(t))^{-\lambda} g(y(t))^{1-\lambda}. \end{aligned}$$

We now have

$$\begin{aligned} \int_{\mathbb{R}} h &= \int_0^1 h(z(t)) z'(t) dt \\ &\geq \int h(\lambda x(t) + (1-\lambda) y(t)) f(x(t))^{-\lambda} g(y(t))^{1-\lambda} dt \\ &\geq 1 \end{aligned}$$

by the assumption on h. This concludes the proof in dimension 1.

Assume now that the Prékopa-Leindler inequality is true in dimension $d-1$, and let's prove it in dimension d. Consider f, g and h satisfying the assumptions of the theorem, and consider their families of $(d-1)$-dimensional restrictions, defined as

$$f_s(x) = f(s, x); \quad s \in \mathbb{R}, \;\; x \in \mathbb{R}^{d-1}$$

and similarly for g and h. It follows from the assumption that

$$h_{\lambda s_0 + (1-\lambda) s_1}(\lambda x + (1-\lambda) y) \geq f_{s_0}(x)^{\lambda} g_{s_1}(y)^{1-\lambda}.$$

Hence

$$\begin{aligned} H(\lambda s_0 + (1-\lambda)s_1) &:= \int_{\mathbb{R}^{d-1}} h_{\lambda s_0 + (1-\lambda)s_1} \\ &\geq \left(\int f_{s_0}\right)^{\lambda} \left(\int g_{s_1}\right)^{1-\lambda} \\ &=: F(s_0)^{\lambda} G(s_1)^{1-\lambda}. \end{aligned}$$

Since F, G and H satisfy the assumptions of the Prékopa-Leindler inequality in dimension one, which we already proved, we can use Fubini's theorem to get

$$\begin{aligned} \int_{\mathbb{R}^d} h &= \int_{\mathbb{R}} H \\ &\geq \left(\int_{\mathbb{R}} F\right)^{\lambda} \left(\int_{\mathbb{R}} G\right)^{1-\lambda} \\ &= \left(\int_{\mathbb{R}^d} f\right)^{\lambda} \left(\int_{\mathbb{R}^d} g\right)^{1-\lambda}. \end{aligned}$$

□

Prékopa-Leindler inequality implies Brunn-Minkowski. Take $e^f = \mathbb{1}_A$, $e^g = \mathbb{1}_B$ and $e^h = \mathbb{1}_{tA+(1-t)B}$, and apply the Prékopa-Leindler inequality.

□

Theorem 14 (Gaussian logarithmic Sobolev inequality)**.** *For all positive, locally lipschitz functions such that* $\int |\nabla f|^2 d\gamma < \infty$*, we have*

$$\int f^2 \log f^2 d\gamma - \left(\int f^2 d\gamma\right) \ln \left(\int f^2 d\gamma\right) \leq 2 \int |\nabla f|^2 d\gamma.$$

Proof. We will prove it in the equivalent form

$$\mathrm{Ent}_{\gamma}(e^g) \leq \frac{1}{2} \int |\nabla g|^2 e^g d\gamma.$$

Let $V(x) = |x|^2/2 + d\ln(2\pi)/2$. We will consider an interpolation between the two densities of interest e^{-V} and e^{g-V} by applying the Prékopa-Leindler inequality with $u(x) = e^{g(x)/t - V(x)}$ and $v(y) = e^{-V(y)}$ for $t \in (0,1)$. The best possible function we can take in the upper bound is $w(z) = e^{g_t(z) - V(z)}$ with

$$g_t(z) = \sup_{z = tx+sy} g(x) - (tV(x) + sV(y) - V(z))$$

where $s = 1 - t$, and we get

$$\int e^{g_t} d\gamma \geq \left(\int e^{g/t} d\gamma\right)^{t}. \tag{2}$$

The entropy arises as the first order-variation of L^p norms as $p \longrightarrow 1$, that is

$$\left(\int e^{(1+\epsilon)g} d\gamma\right)^{1/(1+\epsilon} = \int e^g d\gamma + \epsilon\, \mathrm{Ent}_{\gamma}(e^g) + o(\epsilon),$$

so that replacing $1+\epsilon$ by t^{-1} and letting t go to 1 (i.e. s to zero), we get

$$\left(\int e^{g/t} d\gamma\right)^t = \int e^g d\gamma + s\,\mathrm{Ent}_\gamma(e^g) + o(s).$$

The LSI will be derived by making a Taylor expansion of a suitable upper bound on g_t.

Setting $z = tx + sy$, $w = z - y$ and $r = s/t$, we can rewrite

$$g_t(z) = \sup_w g(z + rw) - \frac{r}{2}|w|^2.$$

Since g has a compact support, for small s (and hence small r) we can write it as

$$g_t(z) = \sup_w g(z) + r\langle \nabla g(z), w\rangle - -\frac{r}{2}|w|^2 + O(r^2|w|^2)$$

where the reminder term is uniformly controlled, so that

for r small enough. Hence, again using that g is compactly supported,

$$\begin{aligned}\int e^{g_t} d\gamma &\leq \int e^g \left(1 + \frac{r}{2}|\nabla g(z)|^2 + C'' r^2\right) d\gamma \\ &= \int e^g d\gamma + \frac{r}{2}\int |\nabla g|^2 e^g d\gamma + O(r^2)\end{aligned}$$

Comparing the two Taylor expansions yields the desired result.

□

4 Lecture 3: Gaussian concentration and isoperimetric inequalities

4.1 Gaussian concentration

Theorem 15 (Gaussian concentration inequality)**.** *Let f be a 1-Lipschitz function on $\mathbb{R}^d$. Then for any $\lambda \in \mathbb{R}$ we have*

$$\int \exp(\lambda f) d\gamma \leq \exp\left(\lambda \int f d\gamma + \frac{\lambda^2}{2}\right).$$

As a consequence, for any $r \geq 0$ and X a standard Gaussian random variable,

$$\mathbb{P}\left(f(X) \geq \mathbb{E}[f(X)] + r\right) \leq \exp(-r^2/2).$$

The first part of the theorem is sharp, since equality is attained for $f(x) = x_1$. The second inequality is not quite sharp for fixed r, but the exponential part is sharp, in the sense that the factor $1/2$ cannot be improved to $1/2+\epsilon$.

For $f(x) = d^{-1/2}\sum x_i$, we observe the same asymptotic as in Cramér's theorem on large deviations, but the result here has the advantage of being non-asymptotic.

We will deduce this inequality from the Gaussian LSI, following what is known as Herbst's argument:

Proof. Let f be a 1-Lipschitz function on $\mathbb{R}^d$ with $\int f d\gamma = 0$, and let $F(\lambda) = \log \int \exp(\lambda f) d\gamma$ for $\lambda \geq 0$. We have $F(0) = 0$ and

$$\begin{aligned} F'(\lambda) &= \frac{\int f \exp(\lambda f) d\gamma}{\int \exp(\lambda f) d\gamma} \\ &= \frac{1}{\lambda} \mathrm{Ent}_\gamma (e^{\lambda f}/e^{F(\lambda)}) + \frac{1}{\lambda} F(\lambda) \\ &\leq \frac{1}{2\lambda} \frac{\int \lambda^2 |\nabla f|^2 \exp(\lambda f) d\gamma}{\int \exp(\lambda f) d\gamma} + \frac{1}{\lambda} F(\lambda) \\ &\leq \frac{\lambda}{2} + \frac{1}{\lambda} F(\lambda). \end{aligned}$$

So we get

$$\left(\frac{1}{\lambda} F \right)' \leq \frac{1}{2}.$$

Integrating with respect to λ yields

$$F(\lambda) \leq \frac{\lambda^2}{2},$$

which is the desired result. The second inequality then follows from Chernoff's inequality (that is, Markov's inequality applied to $e^{\lambda f}$ and then optimizing in λ). □

We will now see an application of Gaussian concentration to data compression. The goal is to find an efficient way of embedding N points of a high-dimensional space Euclidean space (which can be taken as equal to N) in a lower-dimensional space, without distorting too much the distances ebtween points. The result we shall prove is

Theorem 16 (Johnson-Lindenstrauss flattening lemma)**.** *Let $N \in \mathbb{N}$, $\epsilon \in (0, 1)$ and T a set of N points in $\mathbb{R}^N$. Then for any $n > \frac{6 \log(2N^2)}{\epsilon^2}$ tehre exists a linear map $A : \mathbb{R}^N \longrightarrow \mathbb{R}^n$ such that*

$$\forall x, y \in T, \ \ (1 - \epsilon)||x - y||_2 \leq ||Ax - Ay||_2 \leq (1 + \epsilon)||x - y||_2. \tag{3}$$

Proof. The method of proof is probabilistic: we shall construct a linear map at random, and show that (3) holds with positive probability, which ensures existence.

Let $n \in \mathbb{N}$ which shall be chosen later. Let B be a matrix of size $n \times N$ with $B_{i,j} = g_{i,j}$,where the $g_{i,j}$ are iid standard Gaussians. Then for any $u \in \mathbb{R}^N$ such that $||u|| = 1$, Bu is a centered Gaussian vector in $\mathbb{R}^n$, whose covariance matrix is the identity matrix. Since the Euclidean norm is 1-lipschitz, we have

$$\mathbb{P}(|||Bu||_2 - \mathbb{E}[||Bu||_2]| \geq r) = \mathbb{P}(||Bu||_2 - \mathbb{E}[||Bu||] \geq r) + \mathbb{P}(||Bu||_2 - \mathbb{E}[||Bu||] \leq -r) \leq 2\exp(-r^2/2).$$

Let $m = \mathbb{E}[||X||]$ be the expectation of the norm of a standard Gaussian vector in dimension n and $A = \frac{1}{m} B$. With u such that $||u|| = 1$ and taking $r = \epsilon m$, we have

$$\mathbb{P}\left(|||Au|| - 1| \geq \epsilon \right) \leq 2\exp(-\epsilon^2 m^2/2).$$

In particular, for any $x, y \in T$, we have

$$\mathbb{P}\left(\left| \frac{||A(x-y)||}{||x-y||} - 1 \right| \geq \epsilon \right) \leq 2\exp(-\epsilon^2 m^2/2).$$

By a union bound, we get

$$\mathbb{P}\left(\exists x, y \in T \text{ s.t. } \left|\frac{||A(x-y)||}{||x-y||} - 1\right| \geq \epsilon\right) \leq 2N^2 \exp(-\epsilon^2 m^2/2).$$

It is therefore enough to pick n such that

$$m^2 > \frac{2\log(2N^2)}{\epsilon^2}$$

for obtaining existence of a linear map A such that (3) holds.

Since for X standard Gaussian vector we have

$$n = \mathbb{E}[||X||^2] \leq \mathbb{E}[||X||]^{2/3}\mathbb{E}[||X||^4]^{1/3} \text{ and } \mathbb{E}[||X||^4] \leq 3n^2,$$

we have $m^2 \geq n/3$, so it is enough to take $n > \frac{6\log(2N^2)}{\epsilon^2}$ for the result to hold. □

4.2 Isoperimetric inequalities and concentration

The isoperimetric inequality on the sphere, due to P. Lévy and Schmidt, states that

Theorem 17 (Spherical isoperimetric inequality)**.** *Let A be a subset of the unit sphere, and B be a spherical cap with same volume as A. Then*

$$\text{Per}(A) \geq \text{Per}(B).$$

Without loss of generality, we can view a spherical cap as of the form $\{x \in \mathbb{S}^n; 1-t \leq x_1 \leq 1\}$. Its volume is given by the formula

$$\frac{1}{Z_N}\int_{1-t}^{1}(1-x^2)^{N/2-1}dx$$

where Z_N is a constant, computed from the volume ratios.

If we integrate the isoperimetric inequality, we have

Corollary 18. *Let A be a subset of the unit sphere, and B be a spherical cap with same volume as A. Then for any $r > 0$*

$$\text{Vol}(A^r) \geq \text{Vol}(B^r)$$

where A^r is the closed r-neighborhood of A, that is $A^r = \{x \in \mathbb{S}^d, d(x, A) \leq r\}$.

If we consider the uniform probability measure on the unit sphere of dimension d (embedded in $\mathbb{R}^{d+1}$, we can see that the mass of a spherical cap of radius strictly smaller than $\pi/2$ decays exponentially fast. More precisely, for any fixed point z and setting V_d the volume of the d-dimensional unit sphere (which is equal to $\pi^{(d+1)/2}/\Gamma(d/2+1)$, we have

$$\begin{aligned}\frac{\text{Vol}(\{x; d(x,z) \leq t\})}{\text{Vol}(\mathbb{S}^d)} &= \frac{V_{d-1}\int_{\cos(t)}^{1}(1-u^2)^{(d-1)/2}}{V_d} \\ &\leq \frac{\sqrt{2\pi}}{\sqrt{d}}\int_{\cos(t)}^{1}(1-u^2)^{(d-1)/2} \\ &\leq \frac{\sqrt{d+1}}{\sqrt{2\pi}}\int_{\cos(t)}^{1}\exp(-(d-1)u^2/2)du.\end{aligned}$$

We see that as soon as $\pi/2 - t >> d^{-1/2}$, this quantity is exponentially small. So the mass on a high-dimensional sphere is concentrated on an equator.

In particular, if we consider sublevel sets of a lipschitz functions, we see that a 1 lipschitz function of a uniform random variable on the unit sphere has typical fluctuations around its median of order $d^{-1/2}$.

If we rescale the sphere by setting the radius to $\sqrt{n}$ and let n go to infinity, the distribution of a single coordinate converges to a standard Gaussian distribution (Borel-Poincaré lemma). We get in the limit the Gaussian isoperimetric inequality, due to Sudakov, Tsirel'son and Borell. We refer to [18] for a survey of the many known direct proofs of the Gaussian isoperimetric inequality, and some of its extensions.

Theorem 19 (Gaussian isoperimetric inequality)**.** *Let A be a subset of $\mathbb{R}^d$ such that $\gamma_d(A) = \alpha$. Let H be a half-space $]-\infty, a] \times \mathbb{R}^{d-1}$ such that $\gamma_d(H) = \gamma_1(]-\infty, a]) = \alpha$. Then*

$$\mathrm{Per}_{\gamma_d}(A) \geq \frac{\exp(-a^2/2)}{\sqrt{2\pi}} = \mathrm{Per}_{\gamma_d}(H).$$

In this theorem, a can be expressed by inverting the Gaussian cumulative distribution function. Explicit computations show that this inequality is stronger than the Gaussian concentration inequality. More generally, the Gaussian isoperimetric inequality implies the Gaussian logarithmic Sobolev inequality.

5 Lecture 4: Introduction to Stein's method

The goal of this lecture is to introduce the basics about Stein's method for distribution approximation. We refer to [21, 12] for an overview of the field, and to [1, 3] for discussions of recent developments.

5.1 Stein's method for the standard Gaussian measure in dimension one

Definition 20 (L^1 optimal transport distance)**.** *Let μ and ν be two probability measures on a Polish space (E, d), with finite first moment. Then the L^1 Wasserstein (or Monge-Kantorovitch) distance is defined as*

$$W_1(\mu, \nu) := \sup_{f 1-lip} \int f d\mu - \int f d\nu = \inf_{X \equiv \mu, Y \equiv \nu} \mathbb{E}[d(X, Y)].$$

The equality between the two formulations is a non-trivial convex duality result (Kantorovitch-Rubinstein duality theorem).

Theorem 21 (Stein's lemma)**.** *Let $\nu \in \mathcal{P}(\mathbb{R})$. Then*

$$W_1(\nu, \gamma) \leq \sup \left\{ \int (f' - xf) d\nu; \ \ ||f||_\infty, ||f'||_\infty, ||f''||_\infty \leq 1 \right\}$$

Proof. Consider a class of functions $\mathcal{H}$ such that for any 1-lipschitz function g, there exists $f \in \mathcal{H}$ such that $f' - xf = g - \int g d\gamma$. Then trivially

$$W_1(\nu, \gamma) \leq \sup_{f \in \mathcal{H}} \int f' - xf d\nu.$$

So all we need to do is to show that there are solutions that satisfy the regularity bounds. Since the equation is an ODE, this is an explicitly tacklable (although tricky) problem. We refer to [21] for the proof. □

5.2 An application : eigenfunctions of the Laplacian

Our goal here is to prove the following theorem, due to E. Meckes:

Theorem 22. *Let (M, g) be a compact Riemannian manifold with Laplace-Beltrami-operator Δ. Let μ be the normalized volume (i.e. the probability measure proportional to the volume measure induced by g). If h is an eigenfunction of Δ with eigenvalue $-\lambda$, normalized so that $\int h^2 d\mu = 1$, then*

$$W_1(\mu \circ h^{-1}, \gamma) \leq \lambda^{-1} \operatorname{Var}_\mu(|\nabla h|^2).$$

Proof. Let $\nu = \mu \circ h^{-1}$, and let f be a smooth test function on $\mathbb{R}$. We have

$$\begin{aligned}\int xf(x)d\nu &= \int hf \circ hd\mu \\ &= -\lambda^{-1} \int (\Delta_g h) f \circ hd\mu \\ &= \lambda^{-1} \int \nabla h \cdot \nabla (f \circ h) d\mu \\ &= \lambda^{-1} \int f'(h) |\nabla h|^2 d\mu\end{aligned}$$

Hence

$$\int f' - xfd\nu \leq \int f'(h)(1 - \lambda^{-1}|\nabla h|^2)d\mu.$$

Since $\int |\nabla h|^2 d\mu = \lambda$, we consider f such that $||f'||_\infty \leq 1$ and apply Stein's lemma to get

$$W_1(\nu, \gamma) \leq \lambda^{-1} \operatorname{Var}_\mu(|\nabla h|^2).$$

The conclusion immediately follows. □

The argument generalizes to weighted Riemannian manifolds (including the Euclidean space), with reference measure $\mu = e^{-V} d\,\mathrm{Vol}$, if we consider $L = \Delta_g - \nabla V \cdot \nabla$ instead of the usual Laplace-Beltrami operator. The important properties we used in the proof are the integration by parts formula

$$\int (Lf) g e^{-V} d\,\mathrm{Vol} = -\int \nabla f \cdot \nabla g e^{-V} d\,\mathrm{Vol}$$

and the chain rule, which indeed both hold for general reversible diffusion generators.

5.3 The general setting

The key to generalizing this approach to very general measures is a viewpoint proposed by Barbour, known as the generator approach to Stein's method. The starting point is that we can reformulate the caracterization of the Gaussian as

$$\int f'' - xf' d\gamma = 0.$$

The differential operator is $Lf = f'' - xf'$, which is precisely the generator of the Ornstein-Uhlenbeck process

$$dX_t = -X_t dt + \sqrt{2} dB_t.$$

So one way of formulating the characterization is that the standard Gaussian measure is the unique invariant measure of this Markov process.

It is now clear how to generalize the abstract setting: given a target measure μ, one should identify a Markov process with generator L, whose unique invariant probability measure should be μ, and seek an estimate of the form

$$W_1(\mu, \nu) \leq \sup_{f \in \mathcal{F}} \int Lf d\nu$$

where the class of test functions $\mathcal{F}$ should hopefully be as small as possible. To rewrite the Wasserstein distance in this form, we are naturally led to considering the Poisson equation

$$Lf = g - \int g d\mu$$

where g is an arbitrary 1-lipschitz function.

The situation is then about converting good properties of the Markov process into properties of solutions to Poisson equations. Thee are many ways of tackling this problem, and we shall not adress them here.

Let us consider a few examples.

For a multivariate standard Gaussian $\mathcal{N}_d(0, \mathrm{Id})$, it is natural to still consider the Ornstein-Uhlenbeck process, now in dimension d. The generator is

$$Lf = \Delta f - x \cdot \nabla f.$$

The Poisson equation can actually be solved explicitly using a convolution kernel, using properties of Gaussian processes (such as the Ornstein-Uhlenbeck process).

Theorem 23. *Let g be a 1-lipschitz function on $\mathbb{R}^d$. Then there exists a solution f to the PDE*

$$\Delta f - x \cdot \nabla f = g - \int g d\gamma_d$$

such that $||\nabla^2 f||_{HS} \leq 1$.

A remarkable feature is that the estimate is dimension-free, which is very useful in statistical applications.

For a probability measure $\mu = e^{-V} dx$ on $\mathbb{R}^d$, one possible choice is

$$dX_t = \sqrt{2} dB_t - \nabla V(X_t) dt$$

whose generator is $Lf = \Delta f - \nabla V \cdot \nabla f$. The nicer V is, the stronger the properties on the solutions one can prove. Note that the Poisson equation is an elliptic PDE, so nice regularity properties are to be expected (but may be hard to explicitly estimate in a non-compact setting).

For measures on graphs, one can consider random walks, with bias (for example, Metropolis-Hastings algorithm) to enforce a given invariant measure.

6 Introduction to Ricci curvature

The goal here is to give a brief introduction to Ricci curvature, largely following the beginning of [20].

We consider a complete smooth Riemannian manifold (M, g), which for simplicity we assume to be without boundary. g is defined from a family g_x of positive symmetric bilinear forms on the tangent spaces T_xM. In particular, every tangent space is endowed with an Euclidean norm, but this norm changes with the base point.

Recall that the Riemannian structure induces the distance function

$$d(x,y)^2 := \inf \left\{ \int_0^1 g_{u(s)}(\dot{u}(s), \dot{u}(s))ds; \quad u(0) = x; \quad u(1) = y \right\}.$$

Paths that reach the infimum are called geodesics. These play the role in Riemannian manifolds of straight lines in Euclidean spaces. They are constant-speed geodesics if we parametrize them so that $g_u(\dot{u}, \dot{u})$ is constant.

Given a point x and a vector $v \in T_x$, there exists (at least for short times) a unique constant-speed geodesic starting from x with initial velocity v. We denote by $\exp_x(v)$ the exponential map, whose value is the point of M reached by this geodesic at time 1 (which exists if v is small enough).

The Riemannian metric also allows to define a volume measure on M, which is the unique positive measure such that

The Laplace-Beltrami operator is then the operator $\Delta_g f = \mathrm{div}(\nabla f)$. In local coordinate charts, we have the formula

$$\Delta_g f = \frac{1}{\sqrt{det(g_x)}} \sum_{i,j} \partial_i (\sqrt{det(g_x)} (g_x^{-1})_{ij} \partial_j f)$$

It satisfies two important properties we shall repeatedly make use of: the chain rule, and an inetgration by parts formula:

Proposition 24 (Some properties of the Laplace-Beltrami operator). *1. Chain rule :* $\Delta_g \phi(f) = \phi'(f)\Delta_g f + \phi''(f)|\nabla f|^2$

2. Integration by parts: $\int (\Delta_g f) h d\,\mathrm{Vol} = -\int \nabla f \cdot \nabla g d\,\mathrm{Vol}$

In particular, it is a symmetric operator in $L^2(d\,\mathrm{Vol}) \cap \mathrm{Dom}(\Delta)$, *and* $-\Delta_g$ *is nonnegative.*

=

Finally, we now define parallel transport. Let $x, y \in M$ be two distinct points in M. How can we transport a pair of orthogonal vectors u_x in T_xM to T_yM along a geodesic from x to y in a way that is coherent with the geometry? For example, "coherence with the geometry" can be interpreted as preserving orthogonality between pairs of vectors. It can be done in the following way. Let v_x be such that $\exp_x(v_x) = y$, and $w_x \in T_xM$ be a vector orthogonal to v_x. We think of y as being a vector close to x. There exists $w_y \in T_yM$ orthogonal to v_y such that $d(\exp_x(\epsilon w_x), \exp_y(\epsilon w_y))$ is minimized. We may advance to the next point at $\exp_x(\epsilon w_x)$ and perform this process again. We iterate this process while letting ϵ go and $d(x, y)$ go to zero to create parallel transport.

6.1 Sectional and Ricci Curvature

One can view curvature as the evolution of a pair of "parallel" geodesics over time and whether the distance between the two geodesics change over time. To illustrate this intuitive notion, suppose α and β are two distinct parallel geodesics in $\mathbb{R}^n$. In the Euclidean space (zero curvature), the distance $d(\alpha(t), \beta(t))$ between the two paths remains constant over time. In contrast, the distance between two parallel geodesics on a manifold with positive curvature, e.g. $\mathbb{S}^n$, decreases over time. Similarly, the distance between two parallel geodesics on a manifold with negative curvature increases over time.

From this viewpoint (viewing the curvature as an evolution of distance between two geodesic paths), we may consider a Taylor expansion of the distance

$$d(\exp_x(\epsilon w_x), \exp_y(\epsilon w_y)) = \delta\left(1 - \frac{\epsilon^2}{2}K(v, w_x) + O(\epsilon^3 + \delta\epsilon^2)\right). \tag{4}$$

The first order term vanishes since the manifold is locally Euclidean. The **sectional curvature** of the manifold M in the directions (v, w_x) is given by $K(v, w_x)$. From (4), if $K(v, w_x) < 0$, then the distance between the two paths grow larger. On the other hand, if $K(v, w_x) > 0$, then the opposite occur; the distance between the two paths become smaller. Since $K(v, w_x)$ is the 2nd order Taylor coefficient of the distance function, $K(v, w_x)$ can be thought of as how fast curves gets closer or further apart (acceleration of converging/diverging parallel geodesics).

The **Ricci curvature** has the intuitive definition:

$$Ric_x(v) = d \cdot \text{average of } K(v, w) \quad , \quad w \in \mathbb{S}^{n-1} \subset T_xM$$

where d is the manifold dimension. We can also formulate this in terms of the 2nd order Taylor expansion. Let $x \in M$, $y = \exp_x(\delta v)$,

$$S_x = \{\exp_x(\epsilon w) : |w| = 1\} \quad , \quad S_y = \{\exp_y(\epsilon w) : |w| = 1\}.$$

Then we have the average of $d(z_1, z_0)$, $(z_1 \in S_x, z_2 \in S_y)$ along parallel transport. We have the Taylor series expession

$$\text{Average}[d(z_1, z_0)] = \delta\left(1 - \frac{\epsilon^2}{2d}Ric_x(v) + O(\epsilon^3 + \epsilon\delta^2)\right).$$

To obtain the definition of $Ric_x(v, w)$ from $Ric_x(v)$, we can use the polarization identity for quadratic forms. We may interpret the Ricci curvature as follows: $Ric \geq 0$ on average S_x and S_y are closer than $d(x, y)^2$. Another interpretation is given by Sturm and Von Renesse [22]: $Ric \geq 0$ implies that there exists a coupling of Brownian motions such that on average, the distance is non-increasing.

Note that a sphere of dimension d and radius R has constant Ricci curvature, equal to $(d-1)/R^2$.

6.2 Bochner formula

We state a result that characterizes the Ricci curvature by analytic quantities.

Theorem 25 (Bochner's Formula)**.** *For smooth function f, we have:*

$$\frac{1}{2}\Delta_g|\nabla_g f|^2 - \langle\nabla f, \nabla\Delta f\rangle_g = \|\nabla^2 f\|_{HS}^2 + Ric(\nabla f, \nabla f) \tag{5}$$

where $\|A\|_{HS}^2 := \sum_{i,j}(a_{ij})^2$

This formula is a good tool for proving estimates, and can serve as a definition of Ricci curvature, but has the downside of not being very intuitive.

6.3 Some global estimates in positive curvature

Theorem 26 (Bonnet-Myers theorem)**.** *In (M,g), if dim(M) = d, and if the Ricci curvature of M has lower bound $K \cdot g$, where $K > 0$, then the diameter, denoted as Diam(M), is bounded from above by $\pi\sqrt{\frac{d-1}{K}}$.*

This theorem is fairly intuitive: by definition, curvature bounds give local estimates on the distance by the distance on a sphere. This theorem tells us that the same is true globally: the maximal distance between two points on the manifold is bounded from above by the diameter of a sphere with appropriate radius. In particular, spheres are equality cases here.

Theorem 27 (Lichnerowicz theorem)**.** *If $dim(M) = d \geq 2$, $Ric \geq Kg$, $K > 0$; then any positive eigenvalue of $-\Delta$ is greater or equal to $\frac{dK}{d-1}$.*

The sharpest lower bound on positive eigenvalues pf $-\Delta_g$ is known as the spectral gap of the manifold.

Once again, spheres are equality cases in this theorem, which can hence be viewed as a comparison theorem: a manifold with curvature bounded from below by the curvature of a given sphere has spectral gap bounded form below by that of the sphere.

Proof. Let f be an eigenfunction of $-\Delta_g$, so that

$$-\Delta f = \lambda f, \lambda > 0. \tag{6}$$

By the Bochner formula,

$$\frac{1}{2}\Delta|\nabla f|^2 + \lambda|\nabla f| \geq \|\nabla^2 f\|_{HS}^2 + K|\nabla f|^2 \tag{7}$$

By Stokes Theorem,

$$\int_M \Delta|\nabla f|^2 d\,\mathrm{Vol} = 0 \tag{8}$$

Let $\alpha_1, \alpha_2, ..., \alpha_d$ be eigenvalues of $\nabla^2 f$, then:

$$\|\nabla^2 f\|_{HS}^2 = \sum_{i=1}^{d} \alpha_i^2 \geq (\frac{1}{d}\sum \alpha_i)^2 = \frac{1}{d}(\sum \alpha_i)^2 = \frac{1}{d}(\Delta f)^2 = \frac{\lambda^2}{d} f^2 \tag{9}$$

also,

$$\int_M |\nabla f|^2 d\,\mathrm{Vol} = -\int (\Delta f) f d\,\mathrm{Vol} = \lambda \int f^2 d\,\mathrm{Vol} \tag{10}$$

$$\implies \lambda(\lambda - K) \geq \frac{\lambda^2}{d} \tag{11}$$

$$\implies \lambda \geq \frac{dK}{d-1} \tag{12}$$

□

We also have a comparison theorem for isoperimetric profiles:

Theorem 28 (Lévy-Gromov)**.** *If we have $dim(M) = d$, and $\mathrm{Ric} \geq Kg$ everywhere in M, where $K > 0$ and g is the metric of M. Let A be a subset of M, and B be the spherical cap of $S^d(\sqrt{\frac{d-1}{K}})$ so that:*

$$\frac{|A|}{vol(M)} = \frac{|B|}{\mathrm{Vol}(S^d(\sqrt{\frac{d-1}{K}}))} \tag{13}$$

Then, we have $\frac{Per(\partial A)}{\mathrm{Vol}(M)} \geq \frac{Per(B)}{\mathrm{Vol}(S^d(\sqrt{\frac{d-1}{K}}))}$.

Example 29. *In $(S^d(R), g)$, due to the symmetry of geometry of the sphere, $Ricc(S^d(R)) = \frac{d-1}{R^2} \cdot g$ everywhere, with equality holds on the spectral gap(smallest eigenvalue)is equal to $\frac{dk}{d-1}$ and diameter of such sphere is exactly $\pi\sqrt{\frac{d-1}{K}}$, which is the threshold in Bonnet-Myers theorem. The isoperimetric inequality holds true on such sphere.*

Theorem 30 (Obata)**.** *If (M, g) is d-dimensional smooth manifold, $d \geq 2$, $Ric \geq Kg$ with $K > 0$, and either $diam(M) = \pi\sqrt{(d-1)/K}$ or $\lambda_1(-\Delta) = dK/(d-1)$, then M is isometric to the sphere $S^d(R)$ with $R = \sqrt{(d-1)/k}$.*

In the language of optimization, Obata's theorem fully characterizes optimizers for the diameter/spectral gap under a curvature constraint.

6.4 Manifold with General Measures

We want to know what happens to the Bochner formula when (M, g) is endowed with a probability measure of the form $\mu << \mathrm{Vol}$ with $d\mu = e^{-V} d\,\mathrm{Vol}$ (the assumption of unit mass is not so important, statements can be scaled, but it makes sense for applications in probability). We want to consider an operator that preserves the integration by parts formula. It turns out that $L = \Delta_g - \nabla V \cdot \nabla$ is the correct operator to use:

$$\int h \cdot Lf \, d\mu = \int f \cdot Lh \, d\mu = -\int \nabla f \cdot \nabla h \, d\mu.$$

Theorem 31 (Generalized Bochner formula)**.**

$$\frac{1}{2} L|\nabla f|^2 - \langle \nabla f, \nabla(Lf) \rangle = \underbrace{Ric(\nabla f) + \langle \nabla^2 V \nabla f,\ \nabla f \rangle}_{Ric_\mu(\nabla f)} + \|\nabla^2 f\|_{HS}^2$$

This viewpoint goes back to Bakry and Emery's work [4]. The left-hand side is known as the $Gamma_2$ (or *carré du champ*) operator.

Definition 32 (Curvature-dimension condition)**.** *$(M, g, e^{-V}\, d\,\mathrm{Vol})$ satisfies $CD(K, d)$, where K is Ricci curvature of M, if*

$$\frac{1}{2}L|\nabla f|^2 - \langle \nabla f, \nabla(Lf)\rangle \geq K|\nabla f|^2 + \frac{1}{d}(Lf)^2.$$

Recall in the proof of the Lichnerowitz theorem, we used the inequality $\|\nabla^2 f\|_{HS}^2 \geq \frac{1}{d}(\Delta f)^2$. But in a weighted Riemannian manifold, the parameter d does not necessarily stand for the dimension as in the Bochner formula for the unweighted case, it is an algebraic parameter that may take some other value (including negative or infinite).

Theorem 33. *If $CD(K, d)$ holds, then positive eigenvalues of L are greater than or equal to $\dfrac{dK}{d-1}$.*

Examples:

- In $(R^d, \gamma^d = \frac{e^{-|x|^2}}{2\pi^{\frac{d}{2}}})$, we have $Ricci_\mu \geq 1$, and $CD(1, n)$ only holds for $n = \infty$.
- Given $d > 1$, $M = [-1, 1], Lf = (1-x^2)f" - dxf', g = \frac{1}{-x^2+1}, d\mu = \frac{1}{Z^d}(1-x^2)^{\frac{d-2}{2}}dx$, satisfies satisfies $CD(d-1, d)$.

When d is an integer, the second example above corresponds to the first coordinate projection of $S^d(R)$, and Lf is a generator of a Brownian motion in R^d projected down on M. The diameter and eigenvalue on the projected space are the same as for the sphere, so we already see that in the weighted setting, an analogue of Obata's theorem would have to include other cases.

7 Lecture 6: Optimal transport on metric spaces

Definition 34. *A coupling of two probability measures μ and ν, that respectively live on Polish spaces E and F, is a probability measure π on $E \times F$ such that its first marginal is μ and its second marginal is ν. We denote by $\Pi(\mu, \nu)$ the set of all possible couplings of μ and ν.*

Definition 35. *An optimal transport plan between two probability measures μ and ν on a metric space E, with respect to a cost $c : E \times E \longrightarrow \mathbb{R}_+$, is a coupling that minimizes $\int c(x, y)d\pi$ among all possible couplings of μ and ν.*

Definition 36 (Kantorovich problem)**.** *Let E be a Polish space and $\mu, \nu \in \mathbb{P}(E)$. Let $C(\mu, \nu)$ be the set of all couplings satisfying marginal distribution, i.e. $\pi \in C(\mu, \nu)$ satisfies $\pi(A \times E) = \mu(A)$ and $\pi(E \times B) = \nu(B)$. We want to find*

$$\inf_{\pi \in C(\mu,\nu)} E_\pi[c(X, Y)].$$

Definition 37 (Monge problem)**.** *Let E be a Polish space and $\mu, \nu \in \mathbb{P}(E)$. We want to find the transport map T such that the following is minimized:*

$$\inf_{\nu = \mu T^{-1}} E_\mu[c(X, T(X))].$$

Generally, solving the Kantorovich problem is easier than solving the Monge problem. For the Kantorovich problem, we are minimizing a linear function $\int c\,d\pi$ over the convex set $C(\mu,\nu)$, while in Monge problem, the space does not have such a nice structure, and may even be empty.

Definition 38 (Tightness)**.** *We say a set of probability measures $\Gamma \subset \mathcal{P}(E)$ is tight if for $\forall \epsilon > 0$, there exists a compact subset $K_\epsilon subset E$ such that $\inf_{\mu \in Gamma} \mu(K_\epsilon) \geq 1-\epsilon$.*

Theorem 39 (Prokhorov Theorem)**.** *A set of probability measure is closed and tight if and only if it is compact (for the topology of narrow convergence of probability measures).*

Corollary 40. *For any pair of probability measures μ, ν on Polish spaces, $C(\mu,\nu)$ is tight.*

Proof. First, $\forall \epsilon > 0$, given two measures mu and nu, there exist compact subset $K_{1,\epsilon}, K_{2,\epsilon} \in E$ so that $\mu(K_{1,\epsilon}), \nu(K_{2,\epsilon}) \geq 1-\epsilon$. Thus $\forall \pi \in C(\mu,\nu)$, $\pi(K_{1,\epsilon}, K_{2,\epsilon}) \geq 1-2\epsilon$, and hence $C(\mu,\nu)$ is tight. □

It is easy to check that the set of couplings is also closed, so applying Prokhorov's theorem yields compactness.

One can then check that if the cost c is lower-semicontinuous and bounded from below, then the map $\cdot \longrightarrow \int \cdot d\pi$ is also lower-semicontinuous. Together with the compactness of the set of couplings, we conclude that solutions to the Kantorovich problem exist.

Definition 41 (Cyclically Monotone support)**.** *A set $A \subset E \times E$ is c-cyclically monotone if for all $(x_1,y_1),\dots,(x_n,y_n) \in A$*

$$\sum_{i=1}^n c(x_i,y_i) \leq \sum_{i=1}^n c(x_i,y_{i+1}) \quad , \quad n+1 \equiv 1.$$

Proposition 42. *If the cost c is continuous, then the support of an optimal coupling must be c-cyclically monotone.*

The rpoof is by contradiction: if the support is not monotone, then one can build a better coupling by using rearrangements.

7.1 Distance cost

Theorem 43 (Kantorovitch-Rubinstein duality formula)**.** *Let (E,d) be a Polish space, and μ and ν be two probability measures with finite first moment. Then*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y)d\pi = \sup_{f1-lip} \int f d\mu - \int f d\nu.$$

Let $f,g : E \to \mathbb{R}$ be functions where E is a Polish space. Let $c : E \times E \to \mathbb{R}_+$ be a cost function. Consider functions $f(x)+g(y)$ satisfying $f(x)+g(y) \leq c(x,y)$. If $\pi \in C(\mu,\nu)$, then

$$\int c(x,y)\ d\pi \geq \int f(x)+g(y)\ d\pi = \int f(x)d\mu + g(y)\ d\nu.$$

Therefore, we have

$$\inf_{\pi \in C(\mu,\nu)} \int c(x,y)d\pi \geq \sup\{\int f d\mu + g d\nu : f(x)+g(y) \leq c(x,y)\}. \tag{14}$$

Kantorovich duality theorem states that 14 is actually an equality. Meaning, one can express an optimization problem over coupling by an optimization problem over functions:

$$\inf_{\pi \in C(\mu,\nu)} \int c(x,y) d\pi = \sup\{\int f d\mu + g d\nu : f(x) + g(y) \leq c(x,y)\}.$$

We want to make $f(x)+g(y)$ as close as possible to $c(x,y)$. We may then set $g(y) = \inf_x c(x,y) - f(x) = f^c(y)$. This is the best we could do with g, knowing f and c. Similarly, one may set $f(x) = \inf_{y\in E} c(x,y) - g(y) = g^c(x)$. Therefore, we have an optimization over a single function

$$\inf_{\pi \in C(\mu,\nu)} \int c(x,y) d\pi = \sup\{\int f \ d\mu + f^c \ d\nu : f\}.$$

Suppose f maximizes $\sup \int f d\mu + f^c \ d\nu$. We expect $f^{cc} = f$. A function is c-concave if it satisfies this property.

By cyclical monotonicity of optimal couplings, for all $(x,y) \in supp(\pi_{opt})$, we must have $f(x) + f^c(y) = c(x,y)$.

$$\int c(x,y) \ d\pi = \int f(x) \ d\mu + \int f^c(y) \ d\nu \leq \sup_{\tilde{f}} \int \tilde{f} \ d\mu + \int \tilde{f}^c \ d\nu.$$

Since we have inequality in the opposite direction, we have inequality (The Kantorovich Duality Theorem): $\int c(x,y) \ d\pi = \sup_f \int f \ d\mu + \int f^c \ d\nu$.

7.2 Quadratic Cost

Let's consider the Kantorovich duality formulation for the quadratic cost function $c(x,y) = |x-y|^2/2$. Computations show that $|y|^2/2 - f^c(y) = (|x|^2/2 - f(x))^*$ where the $*$ indicate the Legendre transform. From the properties of the Legendre transform (h^* is convex, $h^{**} = h$ iff h is convex), we deduce that the Kantorovich duality formulation for the quadratic cost is given by

$$\inf_{\pi \in C(\mu,\nu)} \int \frac{|x-y|^2}{2} \ d\pi = \sup\{\int \frac{|x|^2}{2} - \varphi \ d\mu + \int \frac{|y|^2}{2} - \varphi^* \ d\nu : \varphi \ \text{ convex}\}.$$

Uisng the differentiability almost everywhere of convex functions, we can show:

Theorem 44 (Brenier-McCann)**.** *If $\mu \in \mathbb{P}(\mathbb{R}^d)$ and μ is absolutely continuous w.r.t. the Lebesgue measure, then there exists a unique convex function φ such that $\mu \circ (\nabla\varphi)^{-1} = \nu$ and*

$$\int |x - \nabla\varphi(x)|^2 \ d\mu = \inf_{\pi \in C(\mu^2)} \int |x-y|^2 \ d\pi.$$

In particular, this implies existence of a solution to the Monge problem, and that solutions to the Monge and Kantorovich problem for the quadratic cost coincide.

7.3 L1 Cost

Now, consider the case when $c(x,y) = |x-y|$. We have $f^c(y) = \inf |x-y| - f(x)$. Suppose a minimizer x_0 exists. If this is the case, then we have $f^c(y_1) - f^c(y_2) \leq |y_1 - y_2|$, i.e. f^c is 1-Lipschitz. Examining the c-transform of a 1-Lipschitz function gives us that if g is 1-Lipschitz, then $g^c = -g$. This implies that the Kantorovich formulation for the L^1 cost is given by

$$\inf_{\pi \in C(\mu,\nu)} \int |x-y| \, d\pi = \sup\{\int f \, d\mu - \int f \, d\nu \ : \ f \in Lip(1)\}.$$

It is possible to prove existence of solutions to the MOnge problem for the distance cost. This was done (after early contributions of Sudakov) with increasing generality in [13, 8, 15, 5].

8 Needle Decompositions and sharp funcitonal inequalities

We will now discuss the following theorem, due to Cavalleti and Mondino [11, 10], following earlier work of Klartag in the Riemannian setting [17].

Theorem 45. *Let (M, g, μ) be a weighted manifold satisfying $CD(K, d)$ with $K \geq 0$, $d < \infty$, $\mu(M) = 1$. If $f \in L^1(M, \mu)$ is such that $E_\mu[f] = 0$, then we can construct the following decomposition:*

1. *$M = Z \cup T$ where $f = 0$ a.e. on Z and $T = \bigsqcup_{q \in Q} N_q$ is a union of disjoint geodesics N_q (except at their endpoints).*
2. *There exists probability measures m_q with $supp(m_q) \subset N_q$ so that (N_q, d, m_q) is a weighted manifold satisfying $CD(K, d)$.*
3. *$\nu \in \mathbb{P}(Q)$ is a probability measure on Q such that for all $B \subset T$ measurable,*

$$\mu(B) = \int_Q m_q(B \cap N_q) \, d\nu(q) \quad \textit{where} \quad \int_{N_q} f \, dm_q = 0.$$

The m_q and ν are obtained by disintegrating the measure μ on the needles N_q.

The needles N_q are constructed by considering the lines connecting points that are in the support of the optimal coupling (for the distance cost) between the two measures with same mass $f_+\mu$ and $f_-\mu$. The fact that they are non-intersecting (except at their endpoints) is a consequence of the cyclical monotonicity of the support. That the average is zero along each needle is a consequence of the mass balancing of the transport (and is delicate to prove rigorously, since the needles are in general of measure zero for μ).

Remark. Two remarks on the above theorem:

1. The above needle decomposition is not proved for $CD(K, \infty)$ since there is no local compactness in infinite dimensional spaces. It is a currently open problem to prove existence of a needle decomposition on locally compact $CD(K, \infty)$ spaces.
2. The converse holds: existence of needle decompositions satisfying curvature bounds is equivalent to the $CD(K, d)$ condition [9].

For each geodesic N_q, we can reparametrize by distance to a point on the geodesic, and we have $(N_q, d) \cong_{iso} ([a,b], |\cdot|)$, $m_q = e^{-\psi}$, satisfying the curvature condition:

$$\frac{L(f')^2}{2} - f' \cdot (Lf)' \geq K(f')^2 + \frac{(Lf)^2}{d} \quad \text{where} \quad Lf = f'' - \psi' f'.$$

Substituting L, we obtain a condition for the density ψ:

$$\psi'' - \frac{(\psi')^2}{d-1} - K \geq 0 \quad , d > 1.$$

If $m_q = h\, dx$, then the $CD(K, d)$ condition gives the condition on the density

$$h^{d/(d-1)} + \frac{k}{d-1}\left(h^{d/(d-1)}\right) \leq 0.$$

As an application of the needle decomposition, we can prove a Poincaré inequality for this weighted manifold. Assume WLOG $E_\mu[f] = 0$. Applying needle decomposition gives us

$$Var_\mu[f] = \int_Q \int_{N_q} f^2 \, dm_q \, d\nu(q).$$

If each N_q satisfies the Poincaré inequality with constant C_P, then

$$Var_\mu[f] = \int_Q \int_{N_q} f^2 \, dm_q \, d\nu(q) \leq C_P \int_Q \int_{N_q} |f'|^2 \, dm_q \, d\nu(q) = C_P \int |\nabla f|^2 \, d\mu.$$

The Poincaré inequality on needles can be proved by ODE methods, using the above parametrization and the differential condition on the density. Hence the needle decomposition allows to deduce statements in arbitrary dimension from a one-dimensional counterpart. The sharp constants can be obtained.

The same approach works for other functional inequalities, such as (log-)Sobolev inequalities and isoperimetric inequalities.

9 Rigid and stable functional inequalities

9.1 Rigidity

We shall make use of Caffarelli's contraction theorem in dimension one

Theorem 46. *If $((a,b), |\cdot|, h\,dx)$ satisfies $CD(1,\infty)$ and $(\log h)'' \leq -1$, then there exits a map $T : \mathbb{R} \to (a,b)$ such that T is 1-Lipschitz and $\gamma T^{-1} = h\,dx$ where γ is a Gaussian measure.*

This theorem holds more generally on Euclidean spaces, but no Riemannian analogue is known at this time. It was proved in [7] and an alternative proof can be found in [14].

If $CD(1,\infty)$ is satisfied, then the Poincaré inequality is satisfied with $C_P \leq 1$. What happens if $C_P = 1$? i.e. there exists f such that $Var_\mu[f] = \|\nabla f\|_2^2$.

Theorem 47 (Cheng and Zhou 2017, Gigli,Ketterer,Kuwada,Ohta 2020)**.** *(M, g, μ) smooth weighted manifold satisfying $CD(1,\infty)$. If there exists $f \neq 0$ such that $\int f^2\, d\mu = \int |\nabla f|^2\, d\mu$, then $M \approx \mathbb{R} \times M'$ and $\mu = \gamma \otimes \mu'$.*

Proof. (Sketch) By needle decomposition, we have

$$\int f^2 \, d\mu = \int \int f^2 \, dm_q \, d\nu \leq \int \int |\nabla f|^2 \, dm_q \, d\nu = \int |\nabla f|^2 \, d\mu. \tag{15}$$

Since both the left hand side and the RHS are equal, equalities must be maintained throughout, i.e. the 2nd inequality above is in fact an equality. Therefore, for almost everywhere, the equality is preserved on each needle. Therefore, following the same reasoning, the below inequality must be an equality:

$$\int f^2 \, dx = \int (f \circ T)^2 \, d\gamma \leq \int (f' \circ T)^2 \, d\gamma = \int (f')^2 h \, dx. \tag{16}$$

This implies that $T' = 1$, i.e. T must be a translation. Therefore, we have an explicit expression of $m_q = \gamma T^{-1}$, i.e. $m_q \sim N(c_q, 1)$ where c_q is the translation. By the equality in the Gaussian Poincaré, we have $f \circ T = x$. Since T is a translation, we have $f(x) = x - c_q$. Note that f is invariant under movement in orthogonal direction w.r.t N_q. So c_q is independent of q.

Take $c_q = 0$. Consider $x \mapsto (f(x), q)$ and define the metric $d(q, q') = \inf\{d(x, x') : x \in N_q, \ x \in N_{q'}\}$. Take $M' = Q$ endowed with the defined distance d. The claim follows. □

Remark. A couple of remarks:

1. The number of Gaussian factors should be the dimension of the first eigenspace of $\Delta - \nabla V \cdot \nabla$.
2. One can get a better sharp Poincaré constant if $CD(d-1, d)$ holds and $diam(M) \leq c < \pi$ (to rule out the sphere). This was originally proved by E. Milman [19], and extended to a non-smooth setting using the needle decomposition by Cavalletti and Mondino.

Analogous results exist for locally Hilbertian $CD(K, d)$ spaces. Equality cases are warped products of spheres. This was originally established by Cheeger and Colding, and generalized by Ketterer.

9.2 Stability

We now discuss partial results on what a space with positive curvature and almost minimal spectral gap looks like.

Theorem 48. *Let (M, g, μ) be a smooth weighted manifold satisfying $CD(1, \infty)$. If $\exists f \neq 0$ such that $(1-\epsilon)\int |\nabla f|^2 \, d\mu \leq Var_\mu[f] \leq \int |\nabla f|^2 \, d\mu$, and normalized so that $\int f^2 d\mu = 1$, then $W_1(\mu \circ f^{-1}, \gamma) << 1$ where γ is the Gaussian measure.*

Proof. (Sketch) Assume $E_\mu[f] = 0$. By the needle decomposition, we have

$$(1-\epsilon) \int \int |f'|^2 \, dm_q \, d\nu \leq \int \int f^2 \, dm_q \, d\nu \leq \int \int |f'|^2 \, dm_q \, d\nu. \tag{17}$$

This implies that in every needle, we have almost equality for the Poincaré inequality. Moreover, on average (out of all needles), the difference between the variance and the L^2 norm of the gradient is small:

$$\int_Q \left(\int_{N_q} |f'|^2 - f^2 \, dm_q \right) d\nu \leq \epsilon \int |\nabla f|^2 \, d\mu. \tag{18}$$

By Markov's inequality, there exists a subset $Q_\epsilon \subset Q$ such that $\forall q \in Q_\epsilon$

$$\int |f'|^2 - f^2 \, dm_q \leq c\sqrt{\epsilon} \int f^2 \, dm_q \quad \text{and} \quad \nu(Q_\epsilon) \geq 1 - c\sqrt{\epsilon}. \tag{19}$$

In each needle, we have

$$(1 - c\sqrt{\epsilon}) \int f^2 \, dm_q = \int f^2 \circ T \, dm_q \leq \int (T')^2 (f' \circ T)^2 \, d\gamma \leq \int |f'|^2 \, dm_q. \tag{20}$$

Therefore, $f \circ T$ is almost the equality case and $T' \approx 1$ in L^2. Take the Hermite polynomial decomposition, $f \circ T$ is almost $x + c$ in $H^1(\gamma)$. Therefore, we have $\int |f' - 1|^2 \, dm_q \lesssim c\sqrt{\epsilon}$. This implies that f looks like x since $f' \approx 1$.

Expanding $Var_{m_q}[f + th] \leq \int |f' + th'|^2 \, dm_q$ yields

$$2t \left(\int f'h' - fh \right) + t^2 \int (h')^2 \geq -c\sqrt{\epsilon} \int |f'|^2. \tag{21}$$

We now consider the class of functions $h \in Lip(1)$ since we want to estimate the Wasserstein distance. We set $t \sim \epsilon^{1/4}$ so that

$$\int h' - xh \, dm_q \approx \int f'h' - fh \, dm_q \leq c\epsilon^{1/4} \quad \text{for all} \quad h \in Lip(1). \tag{22}$$

The leftmost approximation is due to $f' \approx 1$. Applying Stein's lemma yields the desired result. □

Stability of the Poincaré inequality was first considered in the Euclidean setting by De Philippis and Figalli, and revisited using Stein's method by Courtade and Fathi. In the more general RCD setting, it was investigated by Bertrand and Fathi. The sue of the needle decomposition for proving stability was pioneered by Cavalletti and Mondino, starting with other examples (notably, isoperimetric inequalities).

Remark. Couple of remarks for the above theorem:

1. Mai and Ohta (2020) worked on the log-Sobolev and isoperimetric inequality version.
2. The order $\epsilon^{1/4}$ is not sharp. One can get the error to be of order $\epsilon \log(\epsilon)$. The sharp order of magnitude of the error is currently unknown.
3. We believe that there should also be a quantitative statement on how far the metric is from being product, but this is currently an open problem.

We now discuss the $CD(d-1, d)$ case. The first problem is the almost equality case: $d(1 - \epsilon) \int |\nabla f|^2 \, d\mu \leq Var_\mu[f] \leq d \int |\nabla f|^2 \, d\mu$. We may assume that f is an eigenfunction of $Lf = -d(1 + \epsilon)f$. Fathi, Gentil, and Serres proved the Wasserstein estimate: $W_1(\mu \circ f^{-1}, Z_d^{-1}(1-x^2)^{d/2-1} \lesssim c(d)\epsilon$. The exponent in that result is sharp, but the dependence of the prefactor on d is not known.

There is another result by Cavaletti and Mondino which says that there exists x_0 such that

$$\|f - \lambda \cos(d(\cdot, x_0))\|_{L^2(\mu)} \lesssim c(d)\epsilon^{1/(8d+4)}. \tag{23}$$

Cavaletti and Mondino use the needle decomposition to achieve such result. An open problem is to find out what is the sharp exponent, and whether it depends on the dimension d.

Another result is an $\epsilon - \delta$ statement leveraging Gromov's result that says the class of manifolds satisfying $CD(d-1,d)$ is precompact in the Gromov-Hausdorff topology. One then uses a compactness argument to get an $\epsilon - \delta$ statement: For all ϵ there exists δ such that if $C_P \geq d(1-\delta)$, then the Gromov-Hausdorff distance $d_{GH}(M, \text{eq. cases}) \leq \epsilon$. Such statements first appeared in works of Cheeger and Colding.

10 Index of notations

- γ_d stands for the standard Gaussian measure on $\mathbb{R}^d$. If in-context the dimension is unambiguous, we shall simply write it γ.
- $W_p(\mu,\nu)$ is the L^p Wasserstein (or Monge-Kantorovitch) distance between the probability measures μ and ν.
- We denote by $\Pi(\mu,\nu)$ the set of all possible couplings of μ and ν.

References

[1] Andreas Anastasiou, Alessandro Barp, François-Xavier Briol, Bruno Ebner, Robert E. Gaunt, Fatemeh Ghaderinezhad, Jackson Gorham, Arthur Gretton, Christophe Ley, Qiang Liu, Lester Mackey, Chris J. Oates, Gesine Reinert, and Yvik Swan. Stein's method meets computational statistics: a review of some recent developments. *Stat. Sci.*, 38(1):120–139, 2023.

[2] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[3] Ehsan Azmoodeh, Giovanni Peccati, and Xiaochuan Yang. Malliavin-Stein method: a survey of some recent developments. *Mod. Stoch., Theory Appl.*, 8(2):141–177, 2021.

[4] Dominique Bakry and Michel Émery. Diffusions hypercontractives. Sémin. de probabilités XIX, Univ. Strasbourg 1983/84, Proc., Lect. Notes Math. 1123, 177-206 (1985)., 1985.

[5] Stefano Bianchini and Fabio Cavalletti. The Monge problem for distance cost in geodesic spaces. *Commun. Math. Phys.*, 318(3):615–673, 2013.

[6] C. Borell. Convex set functions in *d*-space. *Period. Math. Hungar.*, 6(2):111–136, 1975.

[7] Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Commun. Math. Phys.*, 214(3):547–563, 2000.

[8] Luis A. Caffarelli, Mikhail Feldman, and Robert J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Am. Math. Soc.*, 15(1):1–26, 2002.

[9] Fabio Cavalletti, Nicola Gigli, and Flavia Santarcangelo. Displacement convexity of entropy and the distance cost optimal transportation. *Ann. Fac. Sci. Toulouse Math. (6)*, 30(2):411–427, 2021.

[10] Fabio Cavalletti and Andrea Mondino. Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds. *Invent. Math.*, 208(3):803–849, 2017.

[11] Fabio Cavalletti and Andrea Mondino. Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds. *Geom. Topol.*, 21(1):603–645, 2017.

[12] Sourav Chatterjee. A short survey of Stein's method. In *Proceedings of the International Congress of Mathematicians (ICM 2014), Seoul, Korea, August 13–21, 2014. Vol. IV: Invited lectures*, pages 1–24. Seoul: KM Kyung Moon Sa, 2014.

[13] L. C. Evans and W. Gangbo. *Differential equations methods for the Monge-Kantorovich mass transfer problem*, volume 653 of *Mem. Am. Math. Soc.* Providence, RI: American Mathematical Society (AMS), 1999.

[14] Max Fathi, Nathael Gozlan, and Maxime Prod'homme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calc. Var. Partial Differ. Equ.*, 59(3):18, 2020. Id/No 96.

[15] Mikhail Feldman and Robert J. McCann. Monge's transport problem on a Riemannian manifold. *Trans. Am. Math. Soc.*, 354(4):1667–1697, 2002.

[16] Alessio Figalli and David Jerison. Quantitative stability for the Brunn-Minkowski inequality. *Adv. Math.*, 314:1–47, 2017.

[17] Bo'az Klartag. Needle decompositions in Riemannian geometry. *Mem. Amer. Math. Soc.*, 249(1180):v+77, 2017.

[18] Michel Ledoux. Proofs of the gaussian isoperimetric inequality. *https://perso.math.univ-toulouse.fr/ledoux/files/2022/11/Gaussian-isoperimetry.pdf*, 2017.

[19] Emanuel Milman. Sharp isoperimetric inequalities and model spaces for the curvature-dimension-diameter condition. *J. Eur. Math. Soc. (JEMS)*, 17(5):1041–1078, 2015.

[20] Yann Ollivier. A visual introduction to Riemannian curvatures and some discrete generalizations. In *Analysis and geometry of metric measure spaces. Lecture notes of the 50th Séminaire de Mathématiques Supérieures (SMS), Montréal, Canada, June 27 – July 8, 2011*, pages 197–220. Providence, RI: American Mathematical Society (AMS), 2013.

[21] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[22] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Commun. Pure Appl. Math.*, 58(7):923–940, 2005.