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

Max Fathi

July 5, 2023

1 Introduction

2 Lecture 1: 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 $\lambda \in [0, 1]$.

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 [6] 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 [4] for a proof):

Theorem 11 (Borell's theorem)**.** *If a 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

$$\begin{aligned} g_t(z) &\leq \sup_w g(z) + r\langle \nabla g(z), w\rangle - \frac{r}{2}|w|^2 + Cr^2|w|^2 \\ &\bar{g}(z) + r \sup_w \langle \nabla g(z), w\rangle - \left(\frac{1}{2} - Cr\right)|w|^2 \\ &= g(z) + \frac{r}{2(1-2Cr)}|\nabla g(z)|^2 \\ &\leq g(z) + \frac{r + C'r^2}{2}|\nabla g(z)|^2 \end{aligned}$$

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

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

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

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

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). $\square$

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*

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

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. $\square$

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*

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

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 $r > 0$*

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

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

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

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 [7] 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 [8, 5] 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 [8] 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

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

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

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

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

[5] 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.

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

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

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