Isoperimetry and Convex Geometry in High Dimensions

Boaz Klartag

February 5–6, 2026

These are lecture notes for a mini-course hosted by the Gaspard Monge program for optimization (PGMO) and by the master program in optimization of Paris-Saclay (M2 Optimization).

Contents

Request. Please e-mail me at `boaz.klartag@weizmann.ac.il` with any comments, corrections, or suggestions regarding these lecture notes. In addition, if you are able to produce *tikz* figures similar to (or better than) those drawn on the blackboard in class, I would be glad to include them in the notes.

Lecture 1

The cube and the sphere in high dimensions

In these lectures we study geometry in an n-dimensional Euclidean space when the dimension n is very large, tending to infinity. We will encounter high-dimensional phenomena that do not arise in dimension 3 or 7, say, such as concentration of measure or the emergence of approximately symmetric substructures.

The simplest examples of geometric shapes in $\mathbb{R}^n$ are perhaps the unit cube

$$Q^n = \left[-\frac{1}{2}, \frac{1}{2}\right]^n,$$

and the Euclidean unit sphere

$$S^{n-1} = \{x \in \mathbb{R}^n \, ; \, |x| = 1\},$$

where $|x| = \sqrt{\langle x, x \rangle}$ is the Euclidean norm of the vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and we denote the scalar product of $x, y \in \mathbb{R}^n$ by $\langle x, y \rangle = x \cdot y = \sum_i x_i y_i$. Additional examples of geometric shapes in $\mathbb{R}^n$ include the Euclidean unit ball

$$B^n = \{x \in \mathbb{R}^n \, ; \, |x| \leq 1\},$$

whose features are rather similar to those of the unit sphere, the *cross-polytope* which is the convex hull of the $2n$ vectors

$$\pm e_1, \ldots, \pm e_n \in \mathbb{R}^n,$$

and simplices, where an n-dimensional simplex in $\mathbb{R}^n$ is the convex hull of $n+1$ vectors that affinely span $\mathbb{R}^n$. Here, $e_i \in \mathbb{R}^n$ is the standard i^{th} unit vector. Note that a regular $(n-1)$-dimensional simplex is conveniently represented in $\mathbb{R}^n$ as the convex hull of $e_1, \ldots, e_n \in \mathbb{R}^n$.

1.1 The unit cube

Consider the unit cube $Q^n = [-1/2, 1/2]^n \subseteq \mathbb{R}^n$. There are two relevant lengthscales for this cube: its sidelength, which is 1, and its diameter, which is

$$\sqrt{n} = \left| \left(\frac{1}{2}, \ldots, \frac{1}{2}\right) - \left(-\frac{1}{2}, \ldots, -\frac{1}{2}\right) \right|.$$

Here, the diameter of $K \subseteq \mathbb{R}^n$ is

$$\mathrm{diam}(K) = \sup_{x,y \in K} |x - y|.$$

The $\sqrt{n}$ lengthscale is slightly more prevalent in the analysis of the high-dimensional cube; if we are forced to compare the unit cube to a Euclidean ball of a certain radius, then we should choose a ball of radius on the order of $\sqrt{n}$ (or in some cases $\sqrt{n/\log n}$). For example, what is the typical distance between two random points in the unit cube? That is, let

$$X = (X_1, \dots, X_n) \sim \mathrm{Unif}(Q^n)$$

and

$$Y = (Y_1, \dots, Y_n) \sim \mathrm{Unif}(Q^n)$$

be two independent random vectors, each distributed uniformly in the unit cube Q^n. We are interested in typical values of the random variable $|X - Y|$. Its L^2-norm is easy to compute. Indeed, since $X_1, \dots, X_n, Y_1, \dots, Y_n$ are independent random variables, all distributed uniformly in the interval $[-1/2, 1/2]$, we have

$$\sqrt{\mathbb{E}|X - Y|^2} = \sqrt{\mathbb{E} \sum_{i=1}^{n} (X_i - Y_i)^2} = \sqrt{n \cdot 2\mathrm{Var}(X_1)} = \sqrt{n/6}.$$

The random variable $|X - Y|$ is actually *concentrated* around the value $\sqrt{n/6}$, and in fact, for any $t > 0$, the probability that it deviates from this value by more than t may be bounded as follows:

$$\mathbb{P}\left(\left| |X - Y| - \sqrt{\frac{n}{6}} \right| \geq t \right) \leq C \exp(-ct^2), \tag{1.1}$$

for some universal constants $c, C > 0$. Inequality (1.1) shows that most of the mass of the random vector $X - Y$ is contained in a thin spherical shell of radius $\sqrt{n/6}$ and width $O(1)$. Here $B = O(A)$ means that $|B| \leq CA$, where $C > 0$ is some universal constant. Two sources for such concentration inequalities to be discussed in these lectures are *independence* and *convexity*. Let us describe a proof of (1.1) which relies on statistical independence: Observe that for any $u \in \mathbb{R}$ and $t > 0$,

$$\left| u - \sqrt{\frac{n}{6}} \right| \geq t \qquad \Longrightarrow \qquad \left| u^2 - \frac{n}{6} \right| \geq t \sqrt{\frac{n}{6}}$$

and hence (1.1) would follow once we prove that

$$\mathbb{P}\left(\left| \frac{|X - Y|^2 - \frac{n}{6}}{\sqrt{n}} \right| \geq t \right) \leq C \exp(-ct^2), \tag{1.2}$$

for some universal constants $c, C > 0$. Since the random variable

$$|X - Y|^2 = \sum_{i=1}^{n} (X_i - Y_i)^2$$

is a sum of independent, identically-distributed (i.i.d), bounded random variables, the random variable $|X - Y|^2$ is approximately *a Gaussian random variable* of mean $n/6$ and standard deviation $C\sqrt{n}$. The deviation inequality (1.2) fits with this Gaussian approximation; more precisely, it states that the random variable $|X - Y|^2$ has a uniformly sub-gaussian tail, relative to its mean and variance. This follows from the Bernstein (or Hoeffding) concentration inequality for sums of bounded, independent random variables, which is the subject of a guided exercise below.

Our next question about the cube concerns the volumes of its hyperplane sections. For any $\theta \in S^{n-1}$, writing $\theta^\perp = \{x \in \mathbb{R}^n \, ; \, \langle x, \theta \rangle = 0\}$ for its orthogonal complement, we have

$$1 \leq \mathrm{Vol}_{n-1}\left(\theta^\perp \cap Q^n\right) \leq \sqrt{2} \tag{1.3}$$

where the inequality on the left-hand side is due to Hensley [30] and equality is attained when $\theta = e_i$; there is a stronger version due to Vaaler [54] that follows from the Prékopa-Leindler inequality which will be discussed below. The inequality on the right-hand side of (1.3) is due to Ball [2] (see also the simpler proof in Nazarov and Podkorytov [49]) and equality is attained when $\theta = (1, 1, 0, \ldots, 0)/\sqrt{2}$.

We thus see that volumes of central hyperplane sections of the unit cube can fluctuate between the values 1 and $\sqrt{2}$. What is the "typical value" within this interval $[1, \sqrt{2}]$?

Claim 1.1. *For a typical $\theta \in S^{n-1}$, and in particular for $\theta = (1, \ldots, 1)/\sqrt{n}$, we have*

$$\mathrm{Vol}_{n-1}\left(\theta^\perp \cap Q^n\right) = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{12} \cdot \left(1 + O\left(\frac{1}{n}\right)\right). \tag{1.4}$$

Here, "typical" refers to the uniform probability measure on S^{n-1}, to be described shortly.

Claim 1.1 is related to the classical *Central Limit Theorem* (CLT). Indeed, if

$$X = (X_1, \ldots, X_n) \sim \mathrm{Unif}(Q^n),$$

i.e., the random variables $X_1, \ldots, X_n \sim \mathrm{Unif}([-1/2, 1/2])$ are independent, then

$$\sum_{i=1}^{n} \theta_i X_i = \langle X, \theta \rangle$$

is approximately Gaussian for $\theta = (1, \ldots, 1)/\sqrt{n}$, as well as for other choices of a vector of coefficients $\theta \in S^{n-1}$. More precisely, we have the following classical result:

Theorem 1.2 (CLT, version 1). *For any $\theta \in S^{n-1}$ and $t \in \mathbb{R}$,*

$$\left| \mathbb{P}\left(\sqrt{12}\langle X, \theta \rangle \le t \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2} ds \right| \le C \sum_{i=1}^{n} \theta_i^4, \tag{1.5}$$

where $C > 0$ is a universal constant. (Note that $\sqrt{12}\langle X, \theta \rangle$ is a random variable of mean zero and variance one.)

The usual proof of Theorem 1.2 involves the Fourier inversion formula, see e.g. Feller [22, Chapter XVI] or the guided exercise below.

If $\theta = (1, 0, \dots, 0)$ then $\sum_i \theta_i^4 = 1$ and inequality (1.5) is vacuous. However, for a typical $\theta \in S^{n-1}$, including the case $\theta = (1, \dots, 1)/\sqrt{n}$, we have

$$\sum_i \theta_i^4 = O\left(\frac{1}{n}\right), \tag{1.6}$$

which is the correct rate of approximation in the CLT for the high-dimensional cube.[1]

Let us provide a geometric interpretation of the CLT for the cube. Write $f_\theta : \mathbb{R} \to [0, \infty)$ for the density of the random vector $\langle X, \theta \rangle$. A moment of reflection reveals that

$$f_\theta(t) = \mathrm{Vol}_{n-1}\left(H_{\theta,t} \cap Q^n\right)$$

where

$$H_{\theta,t} = \{x \in \mathbb{R}^n \, ; \, \langle x, \theta \rangle = t\} \tag{1.7}$$

is a hyperplane orthogonal to $\theta \in S^{n-1}$ of distance $|t|$ from the origin. By Fubini's theorem, for $s < t$,

$$\mathrm{Vol}_n\left(\{x \in Q^n \, ; \, s \le \langle x, \theta \rangle \le t\}\right) = \mathbb{P}(s \le \langle X, \theta \rangle \le t) = \int_s^t f_\theta(r) dr.$$

Thus Theorem 1.2 provides Gaussian asymptotic estimates for the volume of the intersection of the unit cube with various planks; a plank is the region in space bounded by two parallel hyperplanes. Observe that

$$\frac{1}{\sqrt{12}} f_\theta\left(\frac{t}{\sqrt{12}}\right) \qquad\qquad (t \in \mathbb{R})$$

is the density of the random variable $\sqrt{12}\langle X, \theta \rangle$. Theorem 1.2 admits the following variant:

[1] It is a better (faster) rate than the $O(1/\sqrt{n})$ rate that we have for the CLT for the *discrete* cube $\{-1, 1\}^n$, and which also appears in the Berry-Esseen bound, see Feller [22, Chapter XVI].

Theorem 1.3 (CLT, version 2). *Under the assumptions of Theorem 1.2,*

$$\left| \frac{1}{\sqrt{12}} f_\theta \left(\frac{t}{\sqrt{12}} \right) - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leq C \sum_{i=1}^{n} \theta_i^4,$$

where $C > 0$ is a universal constant.

Theorem 1.3 with $t = 0$ justifies Claim 1.1 above, and may be used in order to show that the volume of typical central hyperplane sections of the cube concentrate around the value $\sqrt{6/\pi}$. Thus, when considering volumes of hyperplane sections, we observe a simpler behavior for the *high-dimensional* cube than for the cube in dimension 5, say.

Let us also mention that the corresponding question of volumes of *hyperplane projections* of the cube is easier to analyze; for any $\theta \in S^{n-1}$ we have the McMullen formula (see [46]),

$$\mathrm{Vol}_{n-1}\left(Proj_{\theta^\perp}(Q^n)\right) = \sum_{i=1}^{n} |\theta_i|, \tag{1.8}$$

where $Proj_{\theta^\perp} : \mathbb{R}^n \to \theta^\perp$ is the orthogonal projection operator, i.e., $Proj_{\theta^\perp} x = x - \langle x, \theta \rangle \theta$.

1.2 The Euclidean unit ball and sphere

The unit cube in $\mathbb{R}^n$ has volume one. By contrast, the volume of the Euclidean unit ball $B^n = \{x \in \mathbb{R}^n \,;\, |x| \leq 1\}$ is extremely tiny:

$$\kappa_n := \mathrm{Vol}_n(B^n) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} = \left(\frac{\sqrt{2\pi e} + o(1)}{\sqrt{n}} \right)^n. \tag{1.9}$$

This is usually proven along the lines of (1.15) below. We thus need to scale the Euclidean unit ball by a factor of the order of $\sqrt{n}$ in order to obtain a body of volume one. More precisely, the radius of the Euclidean ball of volume one is

$$r_n = \kappa_n^{-1/n} \approx \frac{\sqrt{n}}{\sqrt{2\pi e}},$$

since $\mathrm{Vol}_n(r_n B^n) = r_n^n \cdot \kappa_n = 1$. We scale the Euclidean unit ball by a factor of $\sqrt{n}$, and consider a random vector

$$X \sim \mathrm{Unif}(\sqrt{n} B^n),$$

where $\lambda K = \{\lambda x \,;\, x \in K\}$ for $\lambda \in \mathbb{R}$ and $K \subseteq \mathbb{R}^n$.

Is it true that the random vector $\langle X, \theta\rangle$ is approximately Gaussian for $\theta \in S^{n-1}$, like in the case of the high-dimensional cube?

The answer is yes. In fact, by symmetry, the distribution of $\langle X, \theta\rangle$ does not depend on $\theta \in S^{n-1}$, and we may write $f_\theta(t) = f(t)$ for the density of $\langle X, \theta\rangle$. Thus,

$$f_\theta(t) = \frac{\mathrm{Vol}_{n-1}(H_{\theta,t} \cap \sqrt{n}B^n)}{\mathrm{Vol}_n(\sqrt{n}B^n)} \qquad (t \in \mathbb{R}),$$

with $H_{\theta,t}$ as in (1.7). When $|t| \le \sqrt{n}$, the slice

$$H_{\theta,t} \cap \sqrt{n}B^n$$

is an $(n-1)$-dimensional ball of radius $\sqrt{n-t^2}$, by the Pythagoras theorem. Consequently,

$$f(t) = \frac{\kappa_{n-1}(\sqrt{(n-t^2)_+})^{n-1}}{n^{n/2}\kappa_n} = c_n\left(1 - \frac{t^2}{n}\right)_+^{\frac{n-1}{2}} \tag{1.10}$$

with $c_n = \kappa_{n-1}/(\sqrt{n}\kappa_n) = 1/\sqrt{2\pi} + O(1/n)$ by the Stirling formula. The proof of the CLT for the uniform distribution on the cube requires indirect tools such as the Fourier transform. In contrast, the case of the Euclidean ball is conceptually simpler, even though the random variables $X_1, \dots, X_n$ are no longer independent:[2]

Proposition 1.4. *For any $t \in \mathbb{R}$,*

$$\left| f(t) - \frac{1}{\sqrt{2\pi}}e^{-t^2/2} \right| \le \frac{C}{n}, \tag{1.11}$$

where $C > 0$ is a universal constant.

Proof. If $|t| \ge n^{1/4}$ then $e^{-t^2/2} \le e^{-\sqrt{n}/2} \le C/n$ while

$$\left(1 - \frac{t^2}{n}\right)_+^{\frac{n-1}{2}} \le e^{-t^2(n-1)/(2n)} \le \frac{C}{n},$$

and the bound (1.11) holds true. If $|t| \le n^{1/4}$ then we may use the Taylor approximation $\log(1-x) = -x + O(x^2)$ for $|x| \le 1/2$, which yields

$$\frac{n-1}{2}\log\left(1 - \frac{t^2}{n}\right) = -\frac{n-1}{2}\frac{t^2}{n} + O\left(\frac{t^4}{n}\right) = -\frac{t^2}{2} + O\left(\frac{t^4+1}{n}\right).$$

[2] When was it discovered that the marginals of the high-dimensional sphere are approximately Gaussian? Diaconis and Freedman [18, Section 6] searched in vain for this observation in Poincaré's writings, but found it in Borel's book from 1914 in connection with the kinetic theory of gas.

Therefore, for $|t| \le n^{1/4}$,

$$\left(1 - \frac{t^2}{n}\right)_+^{\frac{n-1}{2}} = \exp\left[-t^2/2 + O\left(\frac{t^4+1}{n}\right)\right] = e^{-\frac{t^2}{2}} + O\left(\frac{1}{n}\right). \tag{1.12}$$

□

Where is the "bulk" of the mass of the high-dimensional Euclidean ball located? One answer is "near the boundary". Recall that a star body in $\mathbb{R}^n$ is a subset $K \subseteq \mathbb{R}^n$ such that $tK \subseteq K$ for $0 \le t \le 1$. A property of the high-dimensional Euclidean ball, or any star body in $\mathbb{R}^n$, is that most of its mass lies near the boundary. Indeed, when $X \sim \mathrm{Unif}(B^n)$, for any $0 \le t \le 1$,

$$\mathbb{P}(|X| \le t) = \frac{\mathrm{Vol}_n(tB^n)}{\mathrm{Vol}_n(B^n)} = t^n. \tag{1.13}$$

It follows that for $n \ge 2$,

$$\mathbb{P}\left(1 - \frac{1}{n} \le |X| \le 1\right) = 1 - \left(1 - \frac{1}{n}\right)^n \ge \frac{1}{2}. \tag{1.14}$$

We see from (1.14) that most of the mass of the unit ball is located at distance only $O(1/n)$ from its boundary, which is the unit sphere. Consequently, the distribution of volume on the high-dimensional Euclidean ball is rather close to that on the high-dimensional sphere, and results on S^{n-1} can often be translated to corresponding results on B^n and vice versa.

Another answer for the above question is that the bulk of the mass of the high-dimensional Euclidean ball (or sphere) is located *near the equator*, as we will now explain.

We slightly prefer to work with the unit sphere S^{n-1}, since it is a *homogeneous space*, admitting a transitive group of symmetries. In other words, all points of the sphere S^{n-1} have an "equal status", while the ball B^n contains "special points" such as the origin. What is the volume of the Euclidean unit sphere? By integrating in polar coordinates,

$$\mathrm{Vol}_n(B^n) = \mathrm{Vol}_{n-1}(S^{n-1}) \cdot \int_0^1 r^{n-1} dr = \frac{1}{n} \cdot \mathrm{Vol}_{n-1}(S^{n-1}).$$

We have thus established the following:

Claim 1.5.

$$\mathrm{Vol}_{n-1}(S^{n-1}) = n\kappa_n.$$

We write σ_{n-1} for the uniform probability measure on S^{n-1}. The probability measure σ_{n-1} can either be viewed as the normalized surface area measure on S^{n-1}, or as the unique rotationally-invariant (Haar) probability measure on S^{n-1}.

It is quite common to replace spherical integrals with a Gaussian computation via integration in polar coordinates. Indeed, let

$$Z = (Z_1, \dots, Z_n)$$

be a standard Gaussian vector in $\mathbb{R}^n$ (i.e., its components are independent, standard Gaussian random variables). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a positively p-homogeneous function (i.e., $f(\lambda x) = \lambda^p f(x)$ for $\lambda > 0$ and $x \in \mathbb{R}^n$) which is locally integrable. Then,

$$\begin{aligned} \mathbb{E} f(Z) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-|x|^2/2} dx \\ &= (2\pi)^{-n/2} \cdot n\kappa_n \cdot \int_0^\infty \int_{S^{n-1}} f(r\theta) e^{-|r\theta|^2/2} r^{n-1} dr d\sigma_{n-1}(\theta) \\ &= C_{n,p} \int_{S^{n-1}} f(\theta) d\sigma_{n-1}(\theta), \end{aligned} \tag{1.15}$$

where $C_{n,p} = 2^{p/2-1} \cdot n \cdot \Gamma(\frac{n+p}{2}) / \Gamma(\frac{n+2}{2})$. For instance, in order to show that a typical vector $\theta \in S^{n-1}$ satisfies (1.6), we may argue as follows:

$$\int_{S^{n-1}} \left(\sum_{i=1}^n \theta_i^4 \right) d\sigma_{n-1}(\theta) = \frac{1}{C_{n,4}} \mathbb{E} \sum_{i=1}^n Z_i^4 = \frac{1}{n(n+2)} \cdot 3n = \frac{3}{n+2},$$

as $\mathbb{E} Z_1^4 = 3$. See the exercises below for more information on the distribution of $\sum_i \theta_i^4$ where $\theta = (\theta_1, \dots, \theta_n)$ is a uniformly-distributed random vector in the sphere S^{n-1}. Another relation between the uniform measures on the ball and the sphere is the following fact, going back to Archimedes in the case $n = 3$.

Proposition 1.6. *For $n \geq 3$, if*

$$X = (X_1, \dots, X_n) \sim \mathit{Unif}(S^{n-1})$$

then

$$(X_1, \dots, X_{n-2}) \sim \mathit{Unif}(B^{n-2}).$$

An analytic way to prove this is to note that by calculus, $(X_1, \dots, X_{n-1})$ has density on B^{n-1} which equals $\tilde{c}_n / \sqrt{1 - |x|^2}$. We now integrate the density of $(X_1, \dots, X_{n-1})$ along a suitable segment and obtain that the density of $(X_1, \dots, X_{n-2})$ at the point $y \in B^{n-2}$ is

$$\tilde{c}_n \int_{-\sqrt{1-|y|^2}}^{\sqrt{1-|y|^2}} \frac{dt}{\sqrt{1 - |y|^2 - t^2}} = \tilde{c}_n \int_{-1}^{1} \frac{ds}{\sqrt{1 - s^2}}$$

which is independent of $y \in B^{n-2}$. Here we changed variables $s = t/\sqrt{1-|y|^2}$.

Corollary 1.7. *If $X = (X_1, \dots, X_n) \sim Unif(S^{n-1})$, then for $t \geq 0$,*

$$\mathbb{P}\left(\sqrt{n}|X_1| \geq t\right) \leq Ce^{-t^2/2} \tag{1.16}$$

where $C > 0$ is a universal constant.

Proof I. Since $(X_1, \dots, X_{n-2}) \sim \text{Unif}(B^{n-2})$, the density of $\sqrt{n}X_1$ equals

$$c_n \left(1 - \frac{x^2}{n}\right)_+^{\frac{n-3}{2}} \qquad (x \in \mathbb{R})$$

with $c_n = 1/\sqrt{2\pi} + O(1/n)$. Hence, for $0 \leq t \leq \sqrt{n}$,

$$\mathbb{P}\left(\sqrt{n}|X_1| \geq t\right) = 2\int_t^{\sqrt{n}} c_n \left(1 - \frac{x^2}{n}\right)_+^{\frac{n-3}{2}} dx \leq C \int_t^{\infty} e^{-x^2(n-3)/(2n)} dx.$$

For $t \notin [1, \sqrt{n}]$ conclusion (1.16) is trivial, while for $1 \leq t \leq \sqrt{n}$ we may use that $\int_t^\infty xe^{-x^2/2} dx = e^{-t^2/2}$ and elementary manipulations to conclude (1.16). □

Proof II (which I essentially learned from Afonso Bandeira). We may assume that $t \leq \sqrt{n}$, since otherwise the probability in question vanishes. Since $\sum_{i=1}^n X_i^2 = 1$ we have

$$\mathbb{P}\left(|X_1| \geq t/\sqrt{n}\right) \leq \mathbb{P}\left(X_1^2 + X_2^2 \geq \frac{t^2}{n}\right) = \mathbb{P}\left(\sum_{i=3}^n X_i^2 \leq 1 - \frac{t^2}{n}\right).$$

The random vector $(X_3, \dots, X_n)$ is distributed uniformly in B^{n-2}, according to Proposition 1.6. Therefore, by (1.13),

$$\mathbb{P}\left(\sum_{i=3}^n X_i^2 \leq 1 - \frac{t^2}{n}\right) = \left(1 - \frac{t^2}{n}\right)^{\frac{n-2}{2}} \leq e^{-\frac{t^2}{2}\cdot\frac{n-2}{n}} \leq Ce^{-\frac{t^2}{2}}$$

with $C = e$. □

In particular, we learn from Corollary 1.7 that when $X \sim \text{Unif}(S^{n-1})$,

$$\mathbb{P}\left(|X_1| \geq 1/10\right) \leq Ce^{-cn},$$

which is exponentially small in the dimension n. Thus,

Proposition 1.8. *Most of the mass of the high-dimensional sphere S^{n-1} is located rather close to the equator*

$$\{x \in S^{n-1}\,;\, x_1 = 0\},$$

i.e., at distance roughly $O(1/\sqrt{n})$ from this equator. By the symmetries of the sphere, the same applies for any equator

$$\{x \in S^{n-1}\,;\, \langle x, \theta\rangle = 0\},$$

with $\theta \in S^{n-1}$.

This startling high-dimensional effect is a manifestation of the *concentration of measure* phenomenon on the high-dimensional sphere.

1.3 The isoperimetric inequality on the sphere

The isoperimetric inequality on the sphere allows us to make effective use of this concentration phenomenon. For $A \subseteq S^{n-1}$ and $\varepsilon > 0$ consider the ε-neighborhood of the set A, defined as

$$A_\varepsilon = \left\{x \in S^{n-1}\,;\, d(x, A) < \varepsilon\right\}$$

where $d(x, A) = \inf_{y\in A} d(x, y)$ and $d(x, y) = |x - y|$ is the Euclidean distance between $x, y \in S^{n-1}$. Another option is to work with the geodesic distance on the sphere, namely $\rho(x, y) = \arccos\langle x, y\rangle \in [0, \pi]$. The Euclidean distance (also called here the "tunnel distance") is always shorter than the geodesic distance, though not by much: it is shorter by a multiplicative factor that does not exceed $\pi/2$. These two distances are essentially equivalent for our needs; note that $\cos\rho(x, y) = 1 - d^2(x, y)/2$.

For example, the ε-neighborhood of the hemisphere

$$H = \left\{x \in S^{n-1}\,;\, x_1 \leq 0\right\},$$

is

$$H_\varepsilon = \left\{x \in S^{n-1}\,;\, x_1 \leq \varepsilon \cdot \sqrt{1 - \varepsilon^2/4}\right\}.$$

Clearly $\sigma_{n-1}(H) = 1/2$, while by the concentration of measure bound (1.16),

$$\sigma_{n-1}(H_\varepsilon) \;=\; \mathbb{P}(X_1 \leq \varepsilon \cdot \sqrt{1 - \varepsilon^2/4}) \;\geq\; \mathbb{P}\left(X_1 \leq \varepsilon/2\right) \;\geq\; 1 - Ce^{-c\varepsilon^2 n}. \qquad (1.17)$$

Thus, the measure of the ε-neighborhood of the hemisphere is very close to one if, say, $\varepsilon = 1/10$ and n is large. The isoperimetric inequality of P. Lévy [25, 52] states that among all sets of σ_{n-1}-measure equal to $1/2$, the hemisphere *minimizes* the measure of the ε-neighborhood. Since the ε-neighborhood of the hemisphere already has relatively large measure, this fact has far-reaching consequences.

Theorem 1.9 (spherical isoperimetric inequality). *For any measurable subset $A \subseteq S^{n-1}$ and any $\varepsilon > 0$,*

$$\sigma_{n-1}(A) \geq \frac{1}{2} \qquad \Longrightarrow \qquad \sigma_{n-1}(A_\varepsilon) \geq \sigma_{n-1}(H_\varepsilon) \tag{1.18}$$

where $H \subseteq S^{n-1}$ is a hemisphere. Moreover, for any $0 < t < 1$,

$$\sigma_{n-1}(A) \geq t \qquad \Longrightarrow \qquad \sigma_{n-1}(A_\varepsilon) \geq \sigma_{n-1}(H_\varepsilon^{(t)})$$

where $H^{(t)} \subseteq S^{n-1}$ is a spherical cap with $\sigma_{n-1}(H^{(t)}) = t$. A spherical cap is the intersection of S^{n-1} with a half-space in $\mathbb{R}^n$.

There are several proofs of the spherical isoperimetric inequality; two symmetrization proofs are explained in Benyamini [5] and in the Appendix of Figiel, Lindenstrauss and Milman [23]. We will discuss the proof of Theorem 1.9 in the next lecture. Thanks to Theorem 1.9 and the bound (1.17), we may leverage the concentration of measure phenomenon as follows:

Corollary 1.10. *For any $A \subseteq S^{n-1}$ and $\varepsilon > 0$,*

$$\sigma_{n-1}(A) \geq \frac{1}{2} \qquad \Longrightarrow \qquad \sigma_{n-1}(A_\varepsilon) \geq 1 - C\exp(-c\varepsilon^2 n), \tag{1.19}$$

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

The constant $1/2$ in (1.19) may be replaced by $1/10$ or any other universal constant, at the expense of adjusting the values of the universal constants C and c. Corollary 1.10 tells us that for any measurable set $A \subseteq S^{n-1}$ with $1/10 \leq \sigma_{n-1}(A) \leq 9/10$, most of the mass of the sphere is located near the boundary of A, i.e., at distance on the order of $O(1/\sqrt{n})$ from the "non-linear equator" ∂A. This provides a rather striking answer to our question: where is the "bulk" of the mass of the high-dimensional sphere located?

While Theorem 1.9 is not too difficult to prove, in these lectures we will only prove the weaker Corollary 1.10, which would suffice for all of our needs. Our proof is based on the classical Brunn-Minkowski inequality from 1887:

Theorem 1.11. *(Brunn-Minkowski) Let $S, T \subseteq \mathbb{R}^n$ be non-empty Borel sets. Then,*

$$\mathrm{Vol}_n(S+T)^{1/n} \geq \mathrm{Vol}_n(S)^{1/n} + \mathrm{Vol}_n(T)^{1/n}, \tag{1.20}$$

where $S + T = \{s + t\,;\, s \in S, t \in T\}$ is the Minkowski sum.[3]

[3] The Minkowski sum of two Borel sets in $\mathbb{R}^n$ is Lebesgue measurable. For more information, see e.g. the first pages in Carleson [15]. Observe that by inner regularity of the Lebesgue measure, the Brunn-Minkowski inequality for Borel sets follows from the corresponding inequality for compact sets.

Note that for any *convex* set $K \subseteq \mathbb{R}^n$ and $r_1, r_2 > 0$,

$$r_1 K + r_2 K = (r_1 + r_2)K. \tag{1.21}$$

In fact, (1.21) is the very definition of a convex set. Therefore, when K is convex,

$$\mathrm{Vol}_n(r_i K)^{1/n} = r_i \cdot \mathrm{Vol}_n(K)^{1/n} \qquad (i = 1, 2)$$

while

$$\mathrm{Vol}_n(r_1 K + r_2 K)^{1/n} = (r_1 + r_2) \cdot \mathrm{Vol}_n(K)^{1/n}.$$

Thus equality holds in the Brunn-Minkowski inequality when S and T are $r_1 K$ and $r_2 K$, respectively. In fact, when S and T are assumed compact, equality in (1.20) holds true if and only if S and T are convex and homothetic, see Henstock and Macbeath [32]. Thus the Brunn-Minkowski inequality is closely related to convex sets, even though convexity does not appear in its formulation. A dimension-free corollary of the Brunn-Minkowski inequality is the following:

Corollary 1.12. *For any Borel sets* $S, T \subseteq \mathbb{R}^n$ *and* $0 < \lambda < 1$,

$$\mathrm{Vol}\left((1-\lambda)S + \lambda T\right) \geq \mathrm{Vol}(S)^{1-\lambda}\mathrm{Vol}(T)^{\lambda}. \tag{1.22}$$

This *multiplicative* Brunn-Minkowski inequality holds true also when S or T are empty, as opposed to Theorem 1.11. In order to prove (1.22), say in the case $\lambda = 1/2$, we apply Theorem 1.11 and the arithmetic/geometric means inequality as follows:

$$\mathrm{Vol}_n\left(\frac{S+T}{2}\right)^{1/n} \geq \frac{\mathrm{Vol}_n(S)^{1/n} + \mathrm{Vol}_n(T)^{1/n}}{2} \geq \left(\sqrt{\mathrm{Vol}_n(S)\mathrm{Vol}_n(T)}\right)^{1/n}.$$

The case of a general $\lambda \in (0, 1)$ is similar.

Proof of Corollary 1.10 using the Brunn-Minkowski inequality. We follow Gromov and Milman [26]. We may assume that $n \geq 3$ and

$$\varepsilon \geq 2/\sqrt{n} \tag{1.23}$$

since otherwise the conclusion is vacuous. Let $A \subseteq S^{n-1}$ satisfy $\sigma_{n-1}(A) \geq 1/2$, and let $B \subseteq S^{n-1}$ be the complement of A_ε. Thus, for $x \in A$ and $y \in B$,

$$|x - y| \geq \varepsilon. \tag{1.24}$$

In order to use the Brunn-Minkowski inequality on volumes in $\mathbb{R}^n$ we need to pass from $(n-1)$-dimensional subsets of the sphere to n-dimensional sets in the unit ball. Fortunately, the uniform probability measure on the sphere is very close to that of the ball. That is, consider the following slight radial extension of the sets A and B into the unit ball:

$$S = \bigcup_{1-\frac{1}{n} \le r \le 1} rA, \qquad T = \bigcup_{1-\frac{1}{n} \le r \le 1} rB.$$

Then,

$$\frac{\mathrm{Vol}_n(S)}{\mathrm{Vol}_n(B^n)} = (1 - (1 - 1/n)^n)\, \sigma_{n-1}(A) \ge \frac{\sigma_{n-1}(A)}{2} \ge \frac{1}{4} \tag{1.25}$$

and similarly

$$\frac{\mathrm{Vol}_n(T)}{\mathrm{Vol}_n(B^n)} \ge \frac{\sigma_{n-1}(B)}{2}. \tag{1.26}$$

By (1.23) and (1.24), for any $x \in S$ and $y \in T$,

$$|x - y| \ge \varepsilon - \frac{2}{n} \ge c\varepsilon$$

for, say, $c = 1/4$. Since x and y are far apart, the *uniform convexity* of the sphere implies that their midpoint is deep inside the ball. That is, for any $x, y \in B^n$ with $|x - y| \ge c\varepsilon$,

$$\left|\frac{x+y}{2}\right|^2 = \frac{|x|^2 + |y|^2}{2} - \frac{|x-y|^2}{4} \le 1 - \tilde{c}\varepsilon^2$$

for some universal constant $c > 0$. Hence,

$$\frac{S+T}{2} \subseteq \sqrt{1 - \tilde{c}\varepsilon^2} \cdot B^n \subseteq \left(1 - \bar{c}\varepsilon^2\right) \cdot B^n.$$

Consequently, from the multiplicative Brunn-Minkowski inequality,

$$(1 - \bar{c}\varepsilon^2)^n \ge \frac{\mathrm{Vol}_n\left(\frac{S+T}{2}\right)}{\mathrm{Vol}_n(B^n)} \ge \frac{\sqrt{\mathrm{Vol}_n(S)\mathrm{Vol}_n(T)}}{\mathrm{Vol}_n(B^n)} \ge \sqrt{\frac{1}{4} \cdot \frac{\sigma_{n-1}(B)}{2}},$$

where we used (1.25) and (1.26) in the last passage. Hence,

$$1 - \sigma_{n-1}(A_\varepsilon) = \sigma_{n-1}(B) \le C(1 - \bar{c}\varepsilon^2)^n \le Ce^{-\bar{c}\varepsilon^2 n},$$

and (1.19) is proven. □

This proof of Corollary 1.10 relies heavily on the *uniform convexity* of the Euclidean ball/sphere, the fact that the midpoint between two points in the ball that are far apart, must lie deep inside the ball. It admits generalization to other uniformly convex sets, see the exercises below.

Hadwiger-Ohman proof of Theorem 1.11. Consider first the case where $S, T \subseteq \mathbb{R}^n$ are two parallel boxes, of edge length $a_1, \dots, a_n > 0$ and $b_1, \dots, b_n > 0$ respectively, we have

$$\mathrm{Vol}_n(S+T) = \prod_{i=1}^{n}(a_i + b_i).$$

Here the boxes may be open or closed; for concreteness let us work here with boxes of the form $\prod_{i=1}^{n}[c_i, d_i)$ where $c_i < d_i$ for all i. The Brunn-Minkowski inequality for two parallel boxes thus amounts to the inequality

$$\left(\prod_{i=1}^{n}(a_i+b_i)\right)^{1/n} \geq \left(\prod_{i=1}^{n} a_i\right)^{1/n} + \left(\prod_{i=1}^{n} b_i\right)^{1/n}.$$

This inequality follows from the arithmetic/geometric means inequality, since

$$\left(\prod_{i=1}^{n}\frac{a_i}{a_i+b_i}\right)^{1/n} + \left(\prod_{i=1}^{n}\frac{b_i}{a_i+b_i}\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^{n}\left[\frac{a_i}{a_i+b_i} + \frac{b_i}{a_i+b_i}\right] = 1. \tag{1.27}$$

We move on to the case of general S and T. By approximation, we may assume that both S and T can be written as the union of finitely many disjoint boxes, all parallel to the axes. Consider representations of S and of T as disjoint unions of finitely many boxes, and write N for the number of boxes appearing in the representation of S plus the number of boxes appearing in the representation of T. We prove (1.20) by induction on N.

Since S and T are non-empty, the base of the induction is the case $N = 2$. In this case, S and T must be two parallel boxes, and the Brunn-Minkowski inequality follows from the arithmetic/geometric means inequality (1.27).

Suppose that $N \geq 3$. Then the representation of the S or of the set T consists of at least two disjoint boxes; without loss of generality assume that it is the set S. Let Q and $\tilde{Q}$ be two disjoint boxes from the representation of S. A crucial observation is that since the boxes Q and $\tilde{Q}$ are disjoint, there exists a hyperplane

$$H \subseteq \mathbb{R}^n$$

parallel to the axes that separates Q from $\tilde{Q}$. Writing H of the form $\{x \in \mathbb{R}^n \, ; \, x_i = t\}$ for some $i = 1, \dots, n$ and $t \in \mathbb{R}$, we look at the half-spaces,

$$H_1 = \{x \in \mathbb{R}^n \, ; \, x_i < t\}, \qquad H_2 = \{x \in \mathbb{R}^n \, ; \, x_i \geq t\}. \tag{1.28}$$

These are two disjoint half-spaces whose union equals $\mathbb{R}^n$. Each of these two halfspaces is disjoint either from the box Q or from the box $\tilde{Q}$. For $i = 1, 2$ denote

$$S_i = S \cap H_i.$$

Clearly, S_i may be represented as a disjoint union of finitely many boxes; in fact, each of the boxes of S contributes at most one box to the representation S_i, with either Q or $\tilde{Q}$ not contributing at all. Thus the total number of disjoint boxes in the representation of S_i is *strictly smaller* than in the representation of S. Set

$$\lambda = \frac{\mathrm{Vol}_n(S_1)}{\mathrm{Vol}_n(S)} \in (0,1), \qquad 1-\lambda = \frac{\mathrm{Vol}_n(S_2)}{\mathrm{Vol}_n(S)}.$$

For $s \in \mathbb{R}$ we consider the hyperplane $\tilde{H} = \tilde{H}(s) = \{x \in \mathbb{R}^n\,;\, x_i = s\}$ that is parallel to H, and we define $\tilde{H}_1 = \tilde{H}_1(s)$ and $\tilde{H}_2 = \tilde{H}_2(s)$ analogously to (1.28), i.e., with t replaced by s. Consider the fraction

$$\frac{\mathrm{Vol}_n(T \cap \tilde{H}_1(s))}{\mathrm{Vol}_n(T)}. \tag{1.29}$$

When we let s vary continuously, the fraction in (1.29) varies continuously from 0 to 1. By the mean value theorem, there exists a hyperplane $\tilde{H}$ parallel to H such that denoting

$$T_i = T \cap \tilde{H}_i \qquad\qquad (i = 1,2)$$

we have

$$\frac{\mathrm{Vol}_n(T_1)}{\mathrm{Vol}_n(T)} = \lambda, \qquad 1-\lambda = \frac{\mathrm{Vol}_n(T_2)}{\mathrm{Vol}_n(T)}.$$

For $i = 1,2$, the set T_i may be represented as a disjoint union of finitely many boxes, where each of the boxes in the representation of T contributes at most one box to the representation T_i. Thus the number of boxes in the representation of T_i is *not larger* than in that of T.

Hence the total number of boxes in the representations of S_i and T_i combined is at most $N-1$. By the induction hypothesis,

$$\mathrm{Vol}_n(S_i + T_i)^{1/n} \geq \mathrm{Vol}_n(S_i)^{1/n} + \mathrm{Vol}_n(T_i)^{1/n}.$$

Observe that the Minkowski sum $S_i + T_i$ is contained in the set $H_i + \tilde{H}_i$, which is a halfspace. Moreover, the halfspace $H_2 + \tilde{H}_2$ is the complement in $\mathbb{R}^n$ to the halfspace $H_1 + \tilde{H}_1$. Consequently $S_1 + T_1$ and $S_2 + T_2$ are two disjoint subsets of $S + T$. Thus

$$\begin{aligned}\mathrm{Vol}_n(S+T) &\geq \sum_{i=1}^{2} \mathrm{Vol}_n(S_i + T_i) \geq \sum_{i=1}^{2} (\mathrm{Vol}_n(S_i)^{1/n} + \mathrm{Vol}_n(T_i)^{1/n})^n \\ &= [\lambda + (1-\lambda)] \left(\mathrm{Vol}_n(S)^{1/n} + \mathrm{Vol}_n(T_i)^{1/n}\right)^n,\end{aligned}$$

completing the proof of (1.20). □

The Brunn-Minkowski inequality implies the isoperimetric inequality in $\mathbb{R}^n$, as we shall now explain. Let $A \subseteq \mathbb{R}^n$ be an open set with a smooth boundary. For $0 < \varepsilon < 1$, the Minkowski sum

$$A + \varepsilon B^n$$

equals the ε-neighborhood of A, which is of course the set

$$A_\varepsilon = \{x \in \mathbb{R}^n \, ; \, d(x, A) < \varepsilon\}$$

where $d(x, A) = \inf_{y \in A} |x - y|$. Assuming that A is bounded and connected, it is proven in multivariate calculus class that

$$\mathrm{Vol}_{n-1}(\partial A) = \lim_{\varepsilon \to 0^+} \frac{\mathrm{Vol}_n(A_\varepsilon) - \mathrm{Vol}_n(A)}{\varepsilon}. \tag{1.30}$$

Corollary 1.13. *For any connected, bounded, open set $A \subseteq \mathbb{R}^n$ with a smooth boundary,*

$$\frac{\mathrm{Vol}_{n-1}(\partial A)}{\mathrm{Vol}_n(A)^{\frac{n-1}{n}}} \geq \frac{\mathrm{Vol}_{n-1}(\partial B)}{\mathrm{Vol}_n(B)^{\frac{n-1}{n}}}, \tag{1.31}$$

where $B \subseteq \mathbb{R}^n$ is any Euclidean ball. Moreover, if $B \subseteq \mathbb{R}^n$ is a ball with $\mathrm{Vol}_n(B) = \mathrm{Vol}_n(A)$ then for any $\varepsilon > 0$,

$$\mathrm{Vol}_n(A_\varepsilon) \geq \mathrm{Vol}_n(B_\varepsilon). \tag{1.32}$$

Proof of Corollary 1.13. We prove (1.32) by the Brunn-Minkowski inequality as follows:

$$\begin{aligned} \mathrm{Vol}_n(A_\varepsilon) &= \mathrm{Vol}_n(A + \varepsilon B^n) \geq \left[\mathrm{Vol}_n(A)^{1/n} + \mathrm{Vol}_n(\varepsilon B^n)^{1/n}\right]^n \\ &= \left[\mathrm{Vol}_n(B)^{1/n} + \varepsilon \mathrm{Vol}_n(B^n)^{1/n}\right]^n = \mathrm{Vol}_n(B + \varepsilon B^n) = \mathrm{Vol}_n(B_\varepsilon), \end{aligned} \tag{1.33}$$

where we used the fact that B is homothetic to B^n and convex, and this yields equality in Brunn-Minkowski.

In order to deduce (1.31), we use that $\mathrm{Vol}_{n-1}(S^{n-1}) = n\mathrm{Vol}_n(B^n)$ and hence

$$\frac{\mathrm{Vol}_{n-1}(\partial B)}{\mathrm{Vol}_n(B)^{\frac{n-1}{n}}} = \frac{\mathrm{Vol}_{n-1}(\partial B^n)}{\mathrm{Vol}_n(B^n)^{\frac{n-1}{n}}} = n\mathrm{Vol}_n(B^n)^{1/n}.$$

Consequently, by (1.33), for any $\varepsilon > 0$,

$$\begin{aligned} \mathrm{Vol}_n(A_\varepsilon) &\geq \left[\mathrm{Vol}_n(A)^{1/n} + \varepsilon \mathrm{Vol}_n(B^n)\right]^n \\ &= \mathrm{Vol}_n(A) + n\varepsilon \mathrm{Vol}_n(A)^{\frac{n-1}{n}} \mathrm{Vol}_n(B^n)^{\frac{1}{n}} + o(\varepsilon) \end{aligned}$$

as $\varepsilon \longrightarrow 0$. Therefore, from formula (1.30) for the surface area,

$$\mathrm{Vol}_{n-1}(\partial A) \geq n \mathrm{Vol}_n(A)^{\frac{n-1}{n}} \mathrm{Vol}_n(B^n)^{\frac{1}{n}} = \mathrm{Vol}_n(A)^{\frac{n-1}{n}} \cdot \frac{\mathrm{Vol}_{n-1}(\partial B)}{\mathrm{Vol}_n(B)^{\frac{n-1}{n}}}.$$

□

Convex sets $K \subseteq \mathbb{R}^n$ come to the world in pairs; this is especially true for *centrally-symmetric* convex sets or convex *cones*. A convex body $K \subseteq \mathbb{R}^n$ is centrally-symmetric when $K = -K$. We recall that the polar body to a convex body $K \subseteq \mathbb{R}^n$ containing the origin in its interior is

$$K^\circ = \{x \in \mathbb{R}^n\, ;\, \forall y \in K,\ \langle x, y \rangle \leq 1\}\,.$$

We have $(K^\circ)^\circ = K$ with $K = K^\circ$ if and only if $K = B^n$ (exercise). When K is a polytope, there is a one-to-one correspondence between the vertices of K and the $(n-1)$-dimensional facets of K°. In particular, the number of vertices of K equals the number of facets of K°.

The bodies K and K° are kind-of "inverses" to each other. For instance, for any invertible, linear map $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$,

$$(T(K))^\circ = (T^{-1})^*(K^\circ).$$

Theorem 1.14 (The Santaló and Bourgain-Milman inequalities). *For any centrally-symmetric convex body $K \subseteq \mathbb{R}^n$,*

$$c^n \mathrm{Vol}_n(B^n)^2 \leq \mathrm{Vol}_n(K)\mathrm{Vol}_n(K^\circ) \leq \mathrm{Vol}_n(B^n)^2, \tag{1.34}$$

where $c > 0$ is a universal constant. In fact, $c = 1/2$ works according to Kuperberg [42].

The left-hand side inequality in (1.34) holds true without the central-symmetry assumption, assuming only that 0 lies in the interior of K. The right-hand side inequality in (1.34) holds true whenever $K \subseteq \mathbb{R}^n$ is a centered convex body, i.e., its barycenter lies at the origin. The Mahler conjecture [44, 45] suggests that $c = 2/\pi$ should work in (1.34), this was proven thus far for $n = 2, 3$, see Iriyeh and Shibata[33]. For a symmetrization proof of the right-hand side of (1.34) see Meyer and Pajor [47], and for a particularly elegant simplification of Kuperberg's proof of the left-hand side of (1.34) see Berndtsson [6].

Exercises.

(1) *Bernstein inequalities* (closely related to Bennett, Hoeffding and Chernoff inequalities; see [56, Chapter 2]): Let $M > 0$ and let $X_1, \ldots, X_n$ be independent random variables. Assume that $\mathbb{E}X_i = 0$ and $\mathbb{P}(|X_i| \le M) = 1$ for all i. We will prove that for all $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge t\sqrt{n}\right) \le Ce^{-c(t/M)^2},$$

where $c, C > 0$ are universal constants. By scaling, we may reduce matters to the case $M = 1$.

(a) We will apply Markov's inequality for exponential moments. Begin by proving that for any $s > 0$,

$$\mathbb{E}e^{sX_1} = \sum_{k=0}^{\infty} \frac{\mathbb{E}(sX_1)^k}{k!} \le e^s - s \le e^{s^2},$$

where the last inequality is obvious for $s > 1$ and follows from $e^s \le 1 + s + s^2 \le s + e^{s^2}$ for $0 < s < 1$.

(b) Given $t > 0$, find an appropriate $s > 0$ so that

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) = \mathbb{P}\left(e^{\sum_{i=1}^n sX_i} \ge e^{st}\right) \le e^{-st}\prod_{i=1}^n \mathbb{E}e^{sX_i} \le e^{-t^2/(4n)}.$$

(2) Recall the proof of (1.9) that you might have learned in your undergraduate studies:

$$\begin{aligned}(2\pi)^{n/2} = \int_{\mathbb{R}^n} e^{-|x|^2/2}dx &= \mathrm{Vol}_{n-1}(S^{n-1}) \cdot \int_0^\infty e^{-r^2/2} r^{n-1} dr \\ &= n\kappa_n \cdot 2^{(n-2)/2}\Gamma(n/2).\end{aligned}$$

(3) Show that c_n from (1.10) satisfies $c_n = 1/\sqrt{2\pi} + O(1/n)$ without using the Stirling formula, but rather by using the Taylor approximation (1.12) as well as the formula

$$c_n^{-1} = \int_{-\sqrt{n}}^{\sqrt{n}} \left(1 - \frac{t^2}{n}\right)^{\frac{n-1}{2}} dt.$$

(4) In this exercise we outline the proof of Theorem 1.3 in the case

$$\theta = \frac{(1, \ldots, 1)}{\sqrt{n}} \in S^{n-1}.$$

(a) Abbreviate $f(t) = f_\theta\left(t/\sqrt{12}\right)/\sqrt{12}$, write $\mathrm{sinc}(x) = \sin(x)/x$ and assume that $n \geq 2$. Use the Fourier inversion formula in order to show that for $t \in \mathbb{R}$,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{sinc}^n\left(\sqrt{\frac{3}{n}}x\right) e^{itx}\,dx.$$

Conclude that

$$\left| f(t) - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leq \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} \left| \mathrm{sinc}^n\left(\sqrt{\frac{3}{n}}x\right) - e^{-x^2/2} \right| dx. \qquad (1.35)$$

(b) We bound the integral in (1.35) by C/n by dividing it into three intervals. Consider first the interval $|x| \leq n^{1/4}$, and use Taylor's theorem in order to show that in this interval, the integrand in (1.35) is at most $C\frac{t^4}{n}e^{-t^2/2}$.

(c) Bound the integral in (1.35) also for $n^{1/4} \leq |x| \leq \sqrt{n}$ and for $|x| \geq \sqrt{n}$ and conclude the proof.

(5) Let $Y \sim \mathrm{Unif}(S^{n-1})$, and let $Z \sim N(0,1)$ be a standard Gaussian. Prove that for any $t \in \mathbb{R}$,

$$\left|\mathbb{P}\left(\sqrt{n}Y_1 \leq t\right) - \mathbb{P}\left(Z \leq t\right)\right| \leq \frac{C}{n}.$$

(6) (*) Let $\Theta = (\Theta_1, \ldots, \Theta_n) \in S^{n-1}$ be a uniformly distributed random vector. Show that

$$\mathbb{P}\left(\sum_{i=1}^{n} \Theta_i^4 \geq \frac{C}{n}\right) \leq \exp(-c\sqrt{n})$$

for some universal constants $C, c > 0$.

(Hint: maybe try to show that $\mathbb{E}\left(\sum_{i=1}^n \Gamma_i^4\right)^p \leq (Cn)^p$ for $p \leq c\sqrt{n}$ and $\Gamma_1, \ldots, \Gamma_n$ being i.i.d standard Gaussians, using that $\mathbb{E}\Gamma_i^{4k} \leq (Ck)^{2k}$).

(7) For a convex polygon $P \subseteq \mathbb{R}^2$ and $t > 0$ and for the unit disc $D = \{x \in \mathbb{R}^2\,;\, |x| < 1\}$, prove that for any $t > 0$,

$$\mathrm{Area}(P + tD) = \mathrm{Area}(P) + t \cdot \mathrm{Length}(\partial P) + \pi t^2.$$

(8) Use the Brunn-Minkowski inequality in order to show that for any convex body $K \subseteq \mathbb{R}^n$ that is centrally-symmetric (i.e. $K = -K$) and any $\theta \in S^{n-1}$, $t \in \mathbb{R}$,

$$\mathrm{Vol}_{n-1}(K \cap \theta^\perp) \geq \mathrm{Vol}_{n-1}(K \cap (t\theta + \theta^\perp)).$$

(9) Let $K \subseteq \mathbb{R}^n$ be a centrally-symmetric convex body (i.e., $K = -K$), and consider the norm $\|x\|_K = \inf\{\lambda \geq 0\,;\, x \in \lambda K\}$ whose unit ball is K. The *modulus of convexity* of K is defined for $0 < \varepsilon < 1$ via

$$\delta(\varepsilon) = \inf\left\{1 - \left\|\frac{x+y}{2}\right\|_K \,;\, x, y \in K, \|x - y\|_K \geq \varepsilon\right\}.$$

(a) Verify that if $K \subseteq \mathbb{R}^n$ is an origin-symmetric ellipsoid, then $\delta(\varepsilon) \geq \varepsilon^2/8$.

(b) Write μ_K for the Lebesgue measure on K, normalized to a probability measure. Prove that for any measurable set $A \subseteq K$ with $\mu_K(A) \geq 1/2$ and any $\varepsilon > 0$,

$$\mu_K(A_\varepsilon) \geq 1 - 2e^{-2n\delta(\varepsilon)},$$

where $A_\varepsilon = \left\{x \in \mathbb{R}^n \,;\, \inf_{y \in A} \|x - y\|_K < \varepsilon\right\}$ is the ε-neighborhood of A with respect to the norm $\|\cdot\|_K$.

Lecture 2

Spherical concentration and the thin shell theorem

In this lecture we discuss applications of the spherical concentration of measure phenomenon in high-dimensions. We begin with the following corollary of Lévy's isoperimetric inequality:

Theorem 2.1 ("spherical concentration of Lipschitz functions"). *Let $f : S^{n-1} \to \mathbb{R}$ be an L-Lipschitz function, i.e., $|f(x) - f(y)| \le L|x - y|$ for all $x, y \in S^{n-1}$. Consider the average of f on the sphere, namely,*

$$E = \int_{S^{n-1}} f \, d\sigma_{n-1}.$$

Then for any $t > 0$,

$$\sigma_{n-1}\left(\left\{x \in S^{n-1} \, ; \, |f(x) - E| \ge t\right\}\right) \le Ce^{-cn(t/L)^2}, \tag{2.1}$$

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

Theorem 2.1 implies that 1-Lipschitz functions on the high-dimensional sphere behave, in certain respects, as if they were nearly constant. Apriori, one might expect such a function to attain values across the entire interval $[0, 1]$, for instance. However, if we sample five random points from the sphere and evaluate a 1-Lipschitz function f at those points, the resulting values will be very close to each other, differing by at most $O(1/\sqrt{n})$.

Proof of Theorem 2.1. We may assume that $L = 1$ (otherwise, replace f by f/L). Abbreviate $\{f \le t\} = \{x \in S^{n-1} \, ; \, f(x) \le t\}$. Let $M \ge 0$ be a median of the function f, i.e.,

$$\sigma_{n-1}(\{f \le M\}) \ge 1/2 \qquad \text{and} \qquad \sigma_{n-1}(\{f \ge M\}) \ge 1/2.$$

(not that it matters, but the median of a continuous function is uniquely determined). Set $A = \{f \le M\}$. Observe that

$$A_t \subseteq \{f \le M + t\},$$

where $A_t = \left\{x \in S^{n-1} \, ; \, \inf_{y \in A} |x - y| < t\right\}$ is the t-neighborhood of A. Since $\sigma_{n-1}(A) \ge 1/2$, by the spherical isoperimetric inequality that we proved in the previous lecture,

$$\sigma_{n-1}(\{f \le M + t\}) \ge \sigma_{n-1}(A_t) \ge 1 - Ce^{-ct^2 n}. \tag{2.2}$$

Similarly, since the t-neighborhood of $\{f \ge M\}$ is contained in $\{f \ge M - t\}$,

$$\sigma_{n-1}(\{f \ge M - t\}) \ge 1 - Ce^{-ct^2 n}. \tag{2.3}$$

From (2.2) and (2.3), for any $t > 0$.

$$\sigma_{n-1}(\{|f - M| \geq t\}) \leq Ce^{-ct^2 n}. \tag{2.4}$$

The expectation of f is rather close to the median. In fact, by (2.4) and Jensen's inequality,

$$\begin{aligned} |E - M| &= \left| \int_{S^{n-1}} f d\sigma_{n-1} - M \right| \leq \int_{S^{n-1}} |f - M| \, d\sigma_{n-1} \\ &= \int_0^\infty \sigma_{n-1}\left(\{|f - M| \geq t\}\right) dt \leq \int_0^\infty Ce^{-ct^2 n} dt \leq \frac{\tilde{C}}{\sqrt{n}}. \end{aligned}$$

This implies that for any $t > 0$,

$$\sigma_{n-1}(\{|f - E| \geq t\}) \leq Ce^{-ct^2 n}. \tag{2.5}$$

Indeed, if $t \leq 1/\sqrt{n}$ then the right-hand side of (2.5) can be assumed at least 1, while if $t \geq 1/\sqrt{n}$, then we may use our bound for $|E - M|$ and note that

$$\{x \in S^{n-1}\, ;\, |f(x) - E| \geq t\} \subseteq \{x \in S^{n-1}\, ;\, |f(x) - M| \geq Ct\}.$$

Now (2.5) follows from (2.4). □

As we see from the proof of Theorem 2.1, we may replace the expectation E in (2.1) by the median M, as well as by other "central values" of f, like the L^2-norm of f when it's non-negative; see the exercise below.

Remark 2.2. Concentration effects go beyond Lipschitz functions, and that it usually suffices to assume that the function f is "Lipschitz on average". For example, the Poincaré inequality on the sphere states that if $f : S^{n-1}$ is a smooth function (or just locally Lipschitz) and $\int_{S^{n-1}} f d\sigma_{n-1} = 0$, then

$$\int_{S^{n-1}} f^2 d\sigma_{n-1} \leq \frac{1}{n-1} \int_{S^{n-1}} |\nabla f|^2 d\sigma_{n-1}. \tag{2.6}$$

Equality holds in (2.6) if and only if $f(x) = x \cdot \theta$ for some $\theta \in \mathbb{R}^n$. This is proven by analyzing spherical harmonics and the spherical Laplacian, see e.g. Müller [48]. There are also L^p-versions of the Poincaré inequality (2.6) where f^2 and $|\nabla f|^2$ are replaced by $|f|^p$ and $|\nabla f|^p$, respectively. A strong and useful inequality is the log-Sobolev inequality on the sphere, see e.g. Bakry, Gentil and Ledoux [1]. As we will see later on, the Poincaré inequality (2.6) implies sub-exponential concentration of Lipschitz functions, which is considerably weaker than the sub-Gaussian concentration of Theorem 2.1.

Our main application of the spherical concentration of measure phenomenon is a version of the "thin-shell theorem" of Sudakov [53] and Diaconis–Freedman [17]. This theorem offers additional insight into why the Gaussian distribution arises in the central limit theorem.

Let $X = (X_1, \dots, X_n)$ be a random vector in $\mathbb{R}^n$ with $\mathbb{E}|X|^2 < \infty$. We say that X is *isotropic* or *normalized* if

$$\mathbb{E}X_i = 0, \quad \mathbb{E}X_i X_j = \delta_{i,j} \qquad \forall i, j = 1, \dots, n,$$

or in short if

$$\mathbb{E}X = 0 \qquad \text{and} \qquad \mathrm{Cov}(X) := \mathbb{E}X \otimes X = \mathrm{Id},$$

where $x \otimes x = (x_i x_j)_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$ for $x \in \mathbb{R}^n$. Equivalently, a random vector X is isotropic if for any $\theta \in S^{n-1}$, the marginal random variable $\langle X, \theta \rangle$ has mean zero and variance one.

Any random vector X satisfying mild conditions can be made isotropic by applying to it an appropriate linear-affine transformation (exercise!). Thus, isotropicity is just a matter of normalization of the random vector; we need to center it and then stretch or shrink it linearly in some orthogonal directions in order too make it balanced in all directions in terms of variance of marginal distributions.

Theorem 2.3 (Thin-shell theorem). *Let X be an isotropic random vector in $\mathbb{R}^n$, and let Z be a real-valued, standard Gaussian random variable. Assume that for some $\varepsilon \geq 0$,*

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \leq \varepsilon^2. \tag{2.7}$$

Then there exists a subset $\mathcal{A} \subseteq S^{n-1}$ with $\sigma_{n-1}(\mathcal{A}) \geq 1 - C\exp(-c\sqrt{n})$, such that for any $\theta \in \mathcal{A}$ and $t \in \mathbb{R}$,

$$|\mathbb{P}(X \cdot \theta \leq t) - \mathbb{P}(Z \leq t)| \leq C\left(\varepsilon^{1/2} + \frac{1}{n^{1/8}}\right), \tag{2.8}$$

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

The exponents $1/2$ and $1/8$ on the right-hand side of (2.8) are non-optimal. Bobkov, Chistyakov and Götze [10, 11] used the Fourier transform as well as other techniques, and essentially obtained $C\varepsilon^2 \log n$ on the right-hand side of (2.8), with a slightly different definition of ε, and with a slightly different probabilistic estimate on θ.

What is the meaning of condition (2.7)? By the Chebyshev–Markov inequality, this condition implies that

$$\mathbb{P}\left(1 - \sqrt{\varepsilon} \leq \frac{|X|}{\sqrt{n}} \leq 1 + \sqrt{\varepsilon}\right) \geq 1 - \varepsilon.$$

Thus, when $\varepsilon \ll 1$, condition (2.7) implies that the bulk of the mass of X is concentrated in a *thin spherical shell*.

Theorem 2.3 tells us that in order to have many approximately Gaussian marginals, it suffices to verify that most of the mass of the random vector X is contained in a thin spherical shell whose width is much smaller than its radius. The fact that the radius must be $\sqrt{n}$ is dictated by the isotropic normalization of X. From the proof of Theorem 2.3 one can see that the thin-shell condition (2.7) is also necessary for the Gaussian approximation phenomenon of the majority of the marginals.

Examples.

(1) Consider the case where $X = (X_1, \dots, X_n)$ and $X_1, \dots, X_n$ are independent random variables with, say, $\mathbb{E}X_i^2 = 1$ and $\mathbb{E}X_i^4 \le 100$ for all i. The thin-shell condition (2.7) holds true with a rather small ε. Indeed, we may compute that

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \le \mathbb{E}\left(\frac{|X|^2}{n} - 1\right)^2 = Var\left(\frac{|X|^2}{n}\right) = \sum_{i=1}^n Var\left(\frac{X_i^2}{n}\right)$$
$$= \frac{1}{n^2}\sum_{i=1}^n \left[\mathbb{E}X_i^4 - 1\right] \le \frac{100}{n}.$$

Thus the standard deviation of $|X|/\sqrt{n}$ is at most $10/\sqrt{n}$, and (2.7) holds true with $\varepsilon = O(n^{-1/2})$, i.e., the width of the thin spherical shell that contains most of the mass of X is only $O(1/\sqrt{n})$ times its radius. Theorem 2.3 thus implies that many of the marginals of X are approximately Gaussian, in accordance with the classical central limit theorem.

(2) Consider a regular simplex circumscribed by the sphere $\sqrt{n}S^{n-1}$. Let X be a discrete random vector in $\mathbb{R}^n$, uniformly distributed on the $n+1$ vertices of this simplex. Note that X is isotropic, and that the mass of X is concentrated in a thin-spherical shell of width $\varepsilon = 0$. Thus, by Theorem 2.3, most of the marginals of X are approximately Gaussian.

(3) Tomorrow we should discuss a recent proof that the uniform distribution on any convex set in $\mathbb{R}^n$, when isotropic, satisfies the requirements of Theorem 2.3 with $\varepsilon = C/\sqrt{n}$, see Klartag and Lehec [40].

(4) A non-example: Let Y be a random vector distributed uniformly on the sphere S^{n-1}, and let τ be a symmetric Bernoulli random variable, independent of Y, i.e., $\mathbb{P}(\tau = 0) = \mathbb{P}(\tau = 1) = 1/2$. Define

$$X = \begin{cases} \sqrt{\frac{n}{2}}\,Y, & \text{if } \tau = 0, \\ \sqrt{\frac{3n}{2}}\,Y, & \text{if } \tau = 1. \end{cases}$$

Observe that X is an isotropic random vector that does *not* satisfy a good thin-shell estimate, since it assigns mass $1/2$ to each of two spheres of very different radii. Consequently, the marginals $\langle X, \theta \rangle$ are all far from Gaussian: each of the two spheres contributes an approximately Gaussian component to the marginal, but their variances are very different. Hence the density of the marginal $X \cdot \theta$ is the average of two Gaussian densities with very different variances, i.e., it is approximately

$$\frac{1}{2}\left[\frac{1}{\sqrt{\pi}}e^{-t^2} + \frac{1}{\sqrt{3\pi}}e^{-t^2/3}\right],$$

which is not close to Gaussian.

The proof of Theorem 2.3 has the following structure: First, we show that a certain observable, defined as a function on the sphere, is concentrated around some unknown value. Then, in order to identify this value, analyze the expectation of the observable. Our observables would be Lipschitz approximations for the functions

$$S^{n-1} \ni \theta \mapsto \mathbb{P}\left(X \cdot \theta \le t\right) \tag{2.9}$$

for $t \in \mathbb{R}$. The function in (2.9) is not necessarily continuous in general, but as we will see it admits good Lipschitz approximations. We begin the proof of Theorem 2.3 with the following:

Lemma 2.4. *Let X and ε be as in Theorem 2.3, and let $Y \sim Unif(S^{n-1})$. Let $f : \mathbb{R} \to \mathbb{R}$ be an L-Lipschitz function. Then there exists a subset $\Theta \subseteq S^{n-1}$ with $\sigma_{n-1}(\Theta) \ge 1 - C\exp(-c\sqrt{n})$ such that for any $\theta \in \Theta$,*

$$\left|\mathbb{E}f(X \cdot \theta) - \mathbb{E}f(\sqrt{n}Y_1)\right| \le CL\left(\frac{1}{n^{1/4}} + \varepsilon\right). \tag{2.10}$$

Proof. For simplicity assume that X has no atom at the origin; it is an exercise to go over the proof below and eliminate this requirement. We may assume that Y is independent of X, since this assumption does not change the values of the various expressions in (2.10). For $\theta \in S^{n-1}$ denote

$$F(\theta) = \mathbb{E}f(X \cdot \theta).$$

Let us observe that F is an L-Lipschitz function on the sphere. Indeed, for any $\theta_1, \theta_2 \in S^{n-1}$,

$$\begin{aligned}|F(\theta_1) - F(\theta_2)| &\le \mathbb{E}\,|f(X\cdot\theta_1) - f(X\cdot\theta_2)| \le L\mathbb{E}\,|X\cdot(\theta_1-\theta_2)| \\ &\le L\sqrt{\mathbb{E}|X\cdot(\theta_1-\theta_2)|^2} = L|\theta_1-\theta_2|,\end{aligned}$$

since X is isotropic, and hence the random variable $X \cdot (\theta_1 - \theta_2)$ has variance $|\theta_1 - \theta_2|^2$. The function F is L-Lipschitz, hence it deviates very little from its average on the sphere. In particular, by using Theorem 2.1 with $t = L/n^{1/4}$, we deduce the existence of a subset $\Theta \subseteq S^{n-1}$ with $\sigma_{n-1}(\Theta) \geq 1 - C\exp(-c\sqrt{n})$ such that

$$\forall \theta \in \Theta, \qquad \left| F(\theta) - \int_{S^{n-1}} F d\sigma_{n-1} \right| \leq \frac{L}{n^{1/4}}. \tag{2.11}$$

The next step is to estimate the average of F on the sphere, and connect it with $\mathbb{E} f(\sqrt{n} Y_1)$. To this end, we observe that the two random variables

$$\langle X/|X|, Y\rangle \qquad \text{and} \qquad Y_1 \tag{2.12}$$

have the same distribution, by the rotational-invariance of the uniform measure on the sphere. Indeed, Y_1 and $\langle Y, \theta\rangle$ have the same distribution for any fixed $\theta \in S^{n-1}$, and the same holds when we replace the fixed $\theta \in S^{n-1}$ by any random vector supported in S^{n-1} that is independent of Y.

Moreover, the random variable $\langle X/|X|, Y\rangle$ is independent of X. Thus, since the two random variables in (2.12) are equidistributed, the same holds when we multiply each of them by $|X|$. It follows that the random variables $\langle X, Y\rangle$ and $|X|Y_1$ are equidistributed. Therefore,

$$\int_{S^{n-1}} F d\sigma_{n-1} = \mathbb{E} F(Y) = \mathbb{E} f(X \cdot Y) = \mathbb{E} f(|X| Y_1). \tag{2.13}$$

Our main assumption (2.7) implies that the random variable $|X|$ is typically very close to $\sqrt{n}$. Thus,

$$\begin{aligned} \left| \mathbb{E} f(|X|Y_1) - \mathbb{E} f(\sqrt{n} Y_1) \right| &\leq L \cdot \mathbb{E} \left| \left(|X| - \sqrt{n} \right) Y_1 \right| \\ &\leq L \sqrt{\mathbb{E} n Y_1^2} \cdot \sqrt{\mathbb{E} \left(|X|/\sqrt{n} - 1 \right)^2} \leq L\varepsilon, \end{aligned}$$

as $\mathbb{E} Y_1^2 = 1/n$. Combining the last inequality with (2.11) and (2.13), the proof is compelte. □

Recall from the first lecture that the density of the random variable $\sqrt{n} Y_1$ is

$$C_n \left(1 - \frac{t^2}{n} \right)^{\frac{n-3}{2}},$$

where $C_n = 1/\sqrt{2\pi} + O(1/n)$, and that if Z is a standard Gaussian random variable then for all $t \in \mathbb{R}$

$$\left| \mathbb{P}\left(\sqrt{n} Y_1 \leq t \right) - \mathbb{P}\left(Z \leq t \right) \right| \leq \frac{C}{n}. \tag{2.14}$$

Proof of Theorem 2.3. Set $\delta = \max\{\sqrt{\varepsilon}, n^{-1/8}\}$. For $t \in \mathbb{R}$ consider the function

$$I_t(x) = \begin{cases} 1 & x < t \\ 1 - (x-t)/\delta & x \in [t, t+\delta] \\ 0 & x > t+\delta \end{cases}$$

Then I_t is a $(1/\delta)$-Lipschitz function, and

$$\mathbb{P}(X \cdot \theta \le t) \ \le\ \mathbb{E} I_t(X \cdot \theta) \ \le\ \mathbb{P}(X \cdot \theta \le t + \delta). \tag{2.15}$$

From Lemma 2.4, for any $t \in \mathbb{R}$ there exists $\mathcal{A}_t \subseteq S^{n-1}$ with

$$\sigma_{n-1}(\mathcal{A}_t) \ge 1 - Ce^{-c\sqrt{n}} \tag{2.16}$$

such that for any $\theta \in \mathcal{A}_t$,

$$\left|\mathbb{E} I_t(X \cdot \theta) \ -\ \mathbb{E} I_t(\sqrt{n} Y_1)\right| \le C \cdot \frac{1}{\delta} \cdot \left(n^{-1/4} + \varepsilon\right) \le \tilde{C}\sqrt{\delta}. \tag{2.17}$$

Our goal is to leverage (2.17) and show that there exists a subset $\mathcal{A} \subseteq S^{n-1}$ of large measure such that for all $\theta \in \mathcal{A}$ and $t \in \mathbb{R}$,

$$|\mathbb{P}(X \cdot \theta \le t) \ -\ \mathbb{P}(Z \le t)| \le C\sqrt{\delta}. \tag{2.18}$$

Step 1. We would like to replace $\sqrt{n} Y_1$ in (2.17) by Z. By the definition of I_t and by (2.14),

$$\begin{aligned} \mathbb{P}(Z \le t) - C/n \le \mathbb{P}\left(\sqrt{n} Y_1 \le t\right) &\le \mathbb{E} I_t(\sqrt{n} Y_1) \\ &\le \mathbb{P}(\sqrt{n} Y_1 \le t + \delta) \le \mathbb{P}(Z \le t + \delta) + \tilde{C}/n. \end{aligned} \tag{2.19}$$

Moreover,

$$|\mathbb{P}(Z \le t + \delta) - \mathbb{P}(Z \le t)| = \int_t^{t+\delta} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \le \frac{1}{\sqrt{2\pi}} \delta \le \delta.$$

Thus by (2.19),

$$\mathbb{E} I_t(\sqrt{n} Y_1) = \mathbb{P}(Z \le t) + O\left(\delta + \frac{1}{n}\right) = \mathbb{P}(Z \le t) + O\left(\delta\right).$$

Consequently, from (2.17), for any $\theta \in \mathcal{A}_t$,

$$|\mathbb{E} I_t(X \cdot \theta) \ -\ \mathbb{P}(Z \le t)| \le C\sqrt{\delta}. \tag{2.20}$$

Step 2. We would need to take care simultaneously of all values of t. To this end, we write

$$\Phi(t) = \mathbb{P}(Z \le t) = \int_{-\infty}^{t} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Consider the Gaussian δ-quantiles

$$t_i = \Phi^{-1}(i \cdot \delta) \qquad \text{for } i = 1, \dots, k := \lceil 1/\delta \rceil \rfloor - 1,$$

so $k \le n^{1/8}$. Then

$$\mathbb{P}(Z \le t_j) = j\delta \qquad \text{for } j = 1, \dots, k.$$

Set also $t_0 = -\infty$ and $t_{k+1} = +\infty$. Consider the event

$$\mathcal{A} = \bigcap_{i=1}^{k} \mathcal{A}_{t_i} \subseteq S^{n-1}$$

which by (2.16) satisfies

$$\sigma_{n-1}(\mathcal{A}) \ge 1 - k \cdot Ce^{-c\sqrt{n}} \ge 1 - Cn^{1/8}e^{-c\sqrt{n}} \ge 1 - \tilde{C}e^{-\tilde{c}\sqrt{n}}.$$

We are now in a position to prove (2.18). Pick $\theta \in \mathcal{A}$ and $t \in \mathbb{R}$. There exists $j = 0, \dots, k$ such that $t_j \le t \le t_{j+1}$. Thus, by (2.20),

$$\begin{aligned}\mathbb{P}(X \cdot \theta \le t) &\le \mathbb{P}(X \cdot \theta \le t_{j+1}) \le \mathbb{E} I_{t_{j+1}}(X \cdot \theta) \le \mathbb{P}(Z \le t_{j+1}) + C\sqrt{\delta} \\ &\le \mathbb{P}(Z \le t) + C\sqrt{\delta} + \delta = \mathbb{P}(Z \le t) + O(\sqrt{\delta}),\end{aligned}$$

which proves one half of the desired inequality (2.18). For the other half, let $i = 0, \dots, k$ be such that $t_i \le t - \delta \le t_{i+1}$. Thus,

$$\begin{aligned}\mathbb{P}(X \cdot \theta \le t) &\ge \mathbb{P}(X \cdot \theta \le t_i + \delta) \ge \mathbb{E} I_{t_i}(X \cdot \theta) \ge \mathbb{P}(Z \le t_i) - C\sqrt{\delta} \\ &\ge \mathbb{P}(Z \le t) - \tilde{C}\sqrt{\delta},\end{aligned}$$

completing the proof of (2.18). □

Exercises.

(1) Let X be a random vector in $\mathbb{R}^n$ with $\mathbb{E}|X|^2 < \infty$ that is not supported by a hyperplane. Prove that there exist a vector $b \in \mathbb{R}^n$ and a positive-definite matrix A such that $A(X) + b$ is isotropic.

(2) Eliminate the requirement that $\mathbb{P}(X = 0) = 0$ from the proof of Lemma 2.4.

(3) Let $(\Omega, \mathbb{P})$ be a probability space, and let $f_1, \dots, f_n \in L^2(\Omega)$ be an orthonormal system such that $\sum_{i=1}^n f_i^2 \equiv 1$, Prove that there exist coefficients $(\theta_1, \dots, \theta_n) \in S^{n-1}$ such that $f = \sum_{i=1}^n \theta_i f_i$ satisfies

$$\left| \mathbb{P}(f \le t) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2} ds \right| \le \frac{C}{n^{1/8}} \qquad (t \in \mathbb{R})$$

where m is the Lebesgue measure. (We remark that there are many non-trivial examples of such orthonormal systems. For instance, any orthonormal basis of the space of spherical harmonics of a certain degree and dimension.)

(4) For a non-negative function $f : S^{n-1} \to \mathbb{R}$, replace E in Theorem 2.1 by $\sqrt{\int_{S^{n-1}} f^2 d\sigma_{n-1}}$.
(jargon: any a with $|a - E| \lesssim L/\sqrt{n}$ may be called a "central value" of f).

Lecture 3

Log-concavity and the Bochner method

Which probability measures in high dimensions enjoy *concentration phenomena*? With respect to which probability measures on $\mathbb{R}^n$ Lipschitz functions are concentrated near their expectation? For which measures on $\mathbb{R}^n$ most of the mass is located near "any equator", and perhaps even "non-linear equators" which are hypersurfaces partitioning space into two parts of equal mass?

Yesterday we considered the case of the uniform measure on the sphere unit S^{n-1}, as well as the closely related uniform measure on the Euclidean unit ball B^n. Furthermore, we know that when

$$X = (X_1, \dots, X_n) \sim \mathrm{Unif}(\sqrt{n} S^{n-1})$$

and n is very large while $k = o(n)$, the random variables

$$X_1, \dots, X_k \in \mathbb{R}^k$$

are approximately independent standard Gaussian random variables in the total variation distance (see Diaconis and Freedman [18, Section 6] for this statement and its history). Thus the standard Gaussian probability measure on $\mathbb{R}^n$ enjoys strong concentration properties, which it inherits from the high-dimensional sphere (see exercise below for a better proof of Gaussian concentration of Lipschitz functions).

There are concentration inequalities available for product measures (i.e., independent random variables), in particular for the boolean cube $\{-1, 1\}^n$, and for random variables with weak dependence properties.

Here we study a class of probability measures in $\mathbb{R}^n$ whose concentration properties were understood relatively recently, which are high-dimensional measures with convexity properties, generalizing uniform distributions on convex sets. In particular, we focus on *log-concave probability measures*.

We begin with the Prékopa-Leindler inequality, which is a functional version of the Brunn-Minkowski inequality.

Theorem 3.1 (Prékopa-Leindler). *Suppose that $f, g, h : \mathbb{R}^n \to [0, \infty)$ are measurable functions and $0 < \lambda < 1$ are such that for any $x, y \in \mathbb{R}^n$,*

$$h\left((1-\lambda)x + \lambda y\right) \geq f(x)^{1-\lambda} g(y)^{\lambda}. \tag{3.1}$$

Then,

$$\int_{\mathbb{R}^n} h \geq \left(\int_{\mathbb{R}^n} f\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g\right)^{\lambda}, \tag{3.2}$$

whenever the integrals on the right-hand side converge.

Remarks.

(1) In the case where $A, B \subseteq \mathbb{R}^n$ have finite volume, by setting

$$f = 1_A, \qquad g = 1_B, \qquad h = 1_{(1-\lambda)A+\lambda B}$$

we recover the Brunn-Minkowski inequality in its multiplicative form. Indeed, f, g, h satisfy the requirements of Theorem 3.1, and hence by its conclusion

$$\begin{aligned} Vol_n\left((1-\lambda)A+\lambda B\right) &= \int h \\ &\geq \left(\int f\right)^{1-\lambda}\left(\int g\right)^{\lambda} = \mathrm{Vol}_n(A)^{1-\lambda}\mathrm{Vol}_n(B)^{\lambda}. \end{aligned}$$

There are also several ways to deduce the Prekopa-Leindler inequality from the Brunn-Minkowski inequality. For example, one may consider convex bodies in higher dimensions whose marginal distributions yield the given functions, and apply Brunn-Minkowski (see, e.g., [35]).

(2) The Prékopa-Leindler inequality may be viewed as a certain converse to Hölder's inequality. Indeed, the Hölder inequality implies that

$$\int_{\mathbb{R}^n} f^{1-\lambda} g^{\lambda} \leq \left(\int_{\mathbb{R}^n} f\right)^{1-\lambda}\left(\int_{\mathbb{R}^n} g\right)^{\lambda}$$

while the Prékopa-Leindler inequality yields

$$\int_{\mathbb{R}^n}\left[\sup_{x=(1-\lambda)y+\lambda z} f(y)^{1-\lambda}g(z)^{\lambda}\right]dx \geq \left(\int_{\mathbb{R}^n} f\right)^{1-\lambda}\left(\int_{\mathbb{R}^n} g\right)^{\lambda}.$$

Proof of Theorem 3.1 *for* $n = 1$. Consider first the case where f and g are bounded functions. If f or g vanish almost everywhere, then there is nothing to prove. Hence we may assume that $\|f\|_\infty$ and $\|g\|_\infty$ are positive numbers. In fact, by homogeneity we may assume that

$$\|f\|_\infty = \|g\|_\infty = 1, \tag{3.3}$$

since otherwise we may replace f by $f/\|f\|_\infty$, replace g by $g/\|g\|_\infty$ and replace h by $h/\left(\|f\|_\infty^{1-\lambda}\|g\|_\infty^{\lambda}\right)$, without affecting the validity of neither the assumptions nor the conclusions of the theorem.

Recall that we abbreviate $\{h > t\} = \{x \in \mathbb{R}\,;\, h(x) > t\}$. Observe that condition (3.1) imply that for all $t > 0$,

$$\{h > t\} \supseteq (1-\lambda)\{f > t\} + \lambda\{g > t\}. \tag{3.4}$$

If $0 < t < 1$ then both sets on the right-hand side of (3.4) are non-empty. Hence, by the one-dimensional Brunn-Minkowski inequality (which is a triviality), for $0 < t < 1$,

$$m(\{h > t\}) \geq (1-\lambda)m(\{f > t\}) + \lambda m(\{g > t\}),$$

where m is the Lebesgue measure on the real line. Therefore,

$$\begin{aligned}\int_{\mathbb{R}} h &= \int_0^\infty m(\{h > t\})dt \geq \int_0^1 m(\{h > t\})dt \\ &\geq (1-\lambda)\int_0^1 m(\{f > t\})dt + \lambda\int_0^1 m(\{g > t\})dt \\ &= (1-\lambda)\int_{\mathbb{R}} f + \lambda\int_{\mathbb{R}} g \geq \left(\int_{\mathbb{R}} f\right)^{1-\lambda}\left(\int_{\mathbb{R}} g\right)^{\lambda}.\end{aligned}$$

This concludes the proof in the case where f and g are bounded. For the general case, for $M > 0$ we replace f by $\min\{f, M\}$, we replace g by $\min\{g, M\}$ and h by $\min\{g, M\}$. Such a truncation still satisfies the requirements of the Prékopa-Leindler inequality (with the same function h). Hence, by the case of the inequality that was already proven,

$$\int_{\mathbb{R}} h \geq \left(\int_{\mathbb{R}} \min\{f, M\}\right)^{1-\lambda}\left(\int_{\mathbb{R}} \min\{g, M\}\right)^{\lambda} \xrightarrow{M\to\infty} \left(\int_{\mathbb{R}} f\right)^{1-\lambda}\left(\int_{\mathbb{R}} g\right)^{\lambda},$$

where we used the monotone convergence theorem in the last passage. □

Proof of Theorem 3.1 for $n \geq 2$. By induction on n. We use $x = (y, t) \in \mathbb{R}^{n-1} \times \mathbb{R}$ as coordinates in $\mathbb{R}^n$ and set

$$\begin{aligned}F(y) &= \int_{-\infty}^\infty f(y,t)\,dt = \int_{-\infty}^\infty f_y(t)\,dt, \\ G(y) &= \int_{-\infty}^\infty g(y,t)\,dt = \int_{-\infty}^\infty g_y(t)\,dt, \\ H(y) &= \int_{-\infty}^\infty h(y,t)\,dt = \int_{-\infty}^\infty h_y(t)\,dt.\end{aligned}$$

We claim that if $y = (1-\lambda)y_1 + \lambda y_2$, for $y_1, y_2 \in \mathbb{R}^n$, then,

$$H(y) \geq F(y_1)^{1-\lambda}G(y_2)^{\lambda}. \tag{3.5}$$

Indeed, if $t = (1-\lambda)t_1 + \lambda t_2$ for $t_1, t_2 \in \mathbb{R}$, then

$$h_y(t) \geq f_{y_1}(t_1)^{1-\lambda}f_{y_2}(t_2)^{\lambda}.$$

Hence (3.5) follows by the one-dimensional Prékopa-Leindler inequality. Thanks to (3.5) and the induction hypothesis, we may apply the $(n-1)$-dimensional Prékopa-Leindler inequality for the functions F, G and H and conclude (3.2). □

Definition 3.2. *A function $\rho : \mathbb{R}^n \to [0, \infty)$ is log-concave if for all $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$,*

$$\rho\left((1-\lambda)x + \lambda y\right) \geq \rho(x)^{1-\lambda}\rho(y)^{\lambda},$$

i.e., if the set $\Omega = \{\rho > 0\}$ is convex and $-\log\rho$ is a convex function on Ω.

We say that a probability measure (or a random vector) in $\mathbb{R}^n$ is log-concave if it is supported in an affine subspace of $\mathbb{R}^n$ with a log-concave density in this subspace. Usually this affine subspace is $\mathbb{R}^n$ itself.

For example, any Gaussian measure in $\mathbb{R}^n$ is log-concave, because its density relative to the affine subspace where it is supported is of the form

$$c_A \exp(-\langle A(x-b), (x-b)\rangle)$$

for a symmetric, positive-definite operator A, a number $C_A > 0$ and a vector $b \in \mathbb{R}^n$. The quadratic function $x \to \langle A(x-b), (x-b)\rangle$ is clearly convex, and hence the Gaussian measure is log-concave. The uniform probability measure on any bounded convex set, is log-concave as well. On the real line, it is very common to encounter log-concave distributions; pretty much, a typical distribution that decays exponentially or faster at infinity is often log-concave. Exponential decay at infinity is indeed a necessary condition for log-concavity. Thus the exponential distribution on $[0, \infty)$ is log-concave, as well as beta and gamma distributions with certain parameters and the double-exponential probability density

$$\exp(-2|x|) \qquad\qquad (x \in \mathbb{R}).$$

Operations that preserve log-concavity include:

(1) Linear images. If X is a log-concave random vector in $\mathbb{R}^n$, then for any subspace $E \subseteq \mathbb{R}^n$ also

$$Proj_E(X)$$

is log-concave, by Prekopa-Leindler. It follows that for any linear (or affine) map $T : \mathbb{R}^n \to \mathbb{R}^m$, the random vector $T(X)$ is log-concave.

(2) Pointwise product. If $f_1, \ldots, f_N$ are log-concave functions, then so is the product $\prod_{i=1}^{N} f_i$. It follows that if a polynomial P has only real roots and is

positive on an interval I, then its restriction to I is log-concave. Indeed,

$$P(x) = c \cdot 1_I(x) \cdot \prod_{i=1}^{N} (x - \lambda_i)$$

for some interval $I \subseteq \mathbb{R}$, a real number $c \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{R} \setminus I$. Since $|x - \lambda_i|$ is log-concave on I, the same applies for P.

(3) Convolution. If $f, g : \mathbb{R}^n \to [0, \infty)$ are log-concave, then $f(y)g(x - y)$ is log-concave on $\mathbb{R}^n \times \mathbb{R}^n$, and consequently the same applies for its marginal $f * g$.

(4) Weak limits. It is an exercise to deduce from the Prékopa-Leindler inequality that if $(\mu_N)_{N \geq 1}$ is a sequence of log-concave concave probability measures converging weakly to a probability measure μ, then μ is also log-concave. This is not an obvious fact; think of the case where μ_N tend to a measure supported on a lower dimensional subspace.

Proposition 3.3 ("How to think on 1D log-concave random variables"). *Let $X \in \mathbb{R}$ be an isotropic, log-concave random variable, i.e., $\mathbb{E}X = 0$ and $Var(X) = 1$. Write ρ for the log-concave density of X. Then for all $x \in \mathbb{R}$,*

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

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

Sketch of proof. For the upper bound, if $\rho(b) < \rho(a)/2$ for some $a < b$, then ρ decays exponentially and in fact $\rho(x) \leq \rho(b) 2^{-x/(b-a)}$ for all $x > b$. As for the lower bound, it is enough to show that $\rho(x) > c'$ for some $x > c''$ and for some $x < -c''$. It is an exercise to filling in the details. □

Corollary 3.4 ("reverse Hölder inequalities", Berwald [7,12]). *For any isotropic, log-concave, real-valued random variable X and for any $p > -1$,*

$$c \cdot \min\{p + 1, 1\} \leq \|X\|_p = (\mathbb{E}|X|^p)^{1/p} \leq C(p + 2), \tag{3.6}$$

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

The case $p = 0$ in (3.6) is interpreted by continuity, i.e.,

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

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

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

Proof of Corollary 3.4. Begin with the inequality on the right-hand side. By the monotonicity of $p \mapsto \|X\|_p$, it is enough to look at $p > 0$. In this case,

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

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

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

and hence

$$\|X\|_p = \left(\mathbb{E}\frac{1}{|X|^q}\right)^{-1/q} \ge (C'(1-q))^{1/q} \ge \tilde{C}(1-q).$$

□

For instance, we learn from Corollary 3.4 that when X is a centered, log-concave random vector in $\mathbb{R}^n$, then

$$\mathbb{E}\langle X, \theta\rangle^4 \le C \left(\mathbb{E}\langle X, \theta\rangle^2\right)^2, \tag{3.7}$$

for a universal constant $C > 0$ (in fact, $C = 9$ is optimal here, see Eitan [19]). Indeed, if $\sigma = (\mathbb{E}\langle X, \theta\rangle^2)^{1/2}$ then $\langle X, \theta\rangle / \sigma$ is an isotropic, log-concave random variable, and (3.7) follows from Corollary 3.4.

Proposition 3.5 ("Reverse Hölder inequalities for polynomials"). *Let X be a real-valued, log-concave random variable, and let $f : \mathbb{R} \to \mathbb{R}$ be a polynomial of degree at most d. Then for any $0 < p \le q$,*

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

for some constant $C_{q,d}$ depending only on q and d.

Proof. Following Bobkov [9], we may assume that f is a monic polynomial in one real variable, hence

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

for some $z_1, \ldots, z_d \in \mathbb{C}$. Consequently, by Hölder's inequality and by Corollary 3.4,

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

□

Remark 3.6. Proposition 3.5 remains valid verbatim if one replaces "real-valued log-concave random variable" by "log-concave random vector in a finite-dimensional normed space"; see Bourgain [13].

Theorem 3.7 (Hensley [31], Fradelizi [24]). *Let $K \subseteq \mathbb{R}^n$ be a centered convex body. Assume that the random vector X that is distributed uniformly in K, is isotropic (or more generally, that $\mathrm{Cov}(X)$ is a scalar matrix). Then for any hyperplanes $H_1, H_2 \subseteq \mathbb{R}^n$ passing through the origin,*

$$Vol_{n-1}(K \cap H_1) \leq C \cdot Vol_{n-1}(K \cap H_2)$$

where $C > 0$ is a universal constant. In fact, $C \leq \sqrt{6}$.

Proof. Let $\theta \in S^{n-1}$ and denote

$$\rho_\theta(t) = \frac{Vol_{n-1}(K \cap (t\theta + \theta^\perp))}{Vol_n(K)}.$$

Then ρ_θ is the density of the random variable $X \cdot \theta$, which is log-concave and isotropic. According to Proposition 3.3, for any $x \in \mathbb{R}$,

$$c' 1_{\{|x| \leq c''\}} \leq \sigma \rho_\theta(x\sigma) \leq Ce^{-c|x|}$$

In particular,

$$c \leq \rho_\theta(0) \leq C,$$

for some universal constants $c, C > 0$. Thus, for $\theta_1, \theta_2 \in S^{n-1}$,

$$\frac{Vol_{n-1}(K \cap \theta_1^\perp)}{Vol_{n-1}(K \cap \theta_2^\perp)} = \frac{\rho_{\theta_1}(0)}{\rho_{\theta_2}(0)} \leq \frac{C}{c} \leq C'.$$

□

Thus, up a multiplicative universal constant, volumes of hyperplane sections of K are closely related to the covariance matrix of the uniform distribution of K.

3.1 Bochner identities and curvature

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

Suppose that μ is an absolutely-continuous, log-concave probability measure in $\mathbb{R}^n$. The measure μ is supported in some open, convex set $K \subseteq \mathbb{R}^n$ (possibly $K = \mathbb{R}^n$), and it has a positive, log-concave density

$$p = e^{-\psi}$$

in K. We will measure distances using the Euclidean metric in $\mathbb{R}^n$, but we will measure volumes using the measure μ. We thus look at the *weighted Riemannian manifold* or the *metric-measure space*

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

We define the Dirichlet energy of a smooth function $f : K \to \mathbb{R}$ as

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

Indeed, we measure the length of the gradient with respect to the Euclidean metric, while we integrate with respect to the measure μ. The *Poincaré constant* of μ, denoted by $C_P(\mu)$, is the minimal number $A > 0$ such that for all μ-integrable, locally-Lipschitz functions $f : K \to \mathbb{R}$ with $\int_K f d\mu = 0$,

$$\int_K f^2 d\mu \leq A \cdot \int_K |\nabla f|^2 d\mu.$$

The Poincaré constant is finite and non-zero (see [8]), and it is a geometric characteristic of the measure μ that is closely related to the isoperimetric inequality. The Poincaré constant of the standard Gaussian measure, for instance, equals one. The inequality

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

where $\mathrm{Var}_\mu(f) = \int f^2 d\mu - (\int f d\mu)^2$, is referred to as the *Poincaré inequality*.

The Laplace-type operator associated with our measure-metric space is defined, initially for $u \in C_c^\infty(K)$, via

$$Lu = L_\mu u = \Delta u - \nabla\psi \cdot \nabla u = e^{\psi} div(e^{-\psi}\nabla u). \tag{3.8}$$

Here, $C_c^\infty(K)$ is the space of smooth functions that are compactly-supported in K. The reason for the definition (3.8) is that for any smooth functions $u, v : \mathbb{R}^n \to \mathbb{R}$, with one of them compactly-supported in K,

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

In particular

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

Thus L is a symmetric operator in $L^2(\mu)$, defined initially for $u \in C_c^\infty(K)$. It can have more than one self-adjoint extension, for example corresponding to the Dirichlet or Neumann boundary conditions when K is bounded.[1]

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

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

$$\varepsilon \cdot \mathrm{Id} \leq \nabla^2\psi(x) \leq \frac{1}{\varepsilon} \cdot \mathrm{Id}. \tag{3.9}$$

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

According to an exercise below, any log-concave probability measure may be approximated arbitrary well by a regular one.

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

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

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

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

[1] When discussing the Bochner technique, it is possible to find ways to circumvent spectral theory of the operator L. Still, spectral theory helps us understand and form intuition, and we will at least quote the relevant spectral theory.

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

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

where the infimum runs over all (say) locally-Lipschitz functions $f \in L^2(\mu)$. Since $\varphi_0 \equiv 1$, we indeed see that the first eigenfunction φ_1 saturates the Poincaré inequality for μ. The linear space

$$\left\{a + Lu\, ;\, a \in \mathbb{R},\ u \in C_c^\infty(\mathbb{R}^n)\right\}$$

is dense in $L^2(\mu)$. For proofs of these spectral theoretic facts, see references in [38].

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

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

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

Proposition 3.8 (Integrated Bochner's formula). *For any $u \in C_c^\infty(\mathbb{R}^n)$,*

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

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

Proof. Integration by parts gives

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

□

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

Theorem 3.9 (improved log-concave Lichnerowicz inequality). *Let $t > 0$ and assume that $\nabla^2\psi(x) \geq t$ for all $x \in \mathbb{R}^n$. Then,*

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

where $\|A\|_{op}$ is the operator norm of the symmetric matrix $A \in \mathbb{R}^{n\times n}$.

Equality in Theorem 3.9 is attained when μ is a Gaussian measure, with any covariance matrix.

Proof of Theorem 3.9. Denote $f = \varphi_1$, the first eigenfunction, normalized so that

$$\|f\|_{L^2(\mu)} = 1.$$

Set $\lambda = 1/C_P(\mu)$. By the Bochner formula and the Poincaré inequality for $\partial^i f$ ($i = 1, \dots, n$),

$$\begin{aligned}
\lambda^2 &= \int_{\mathbb{R}^n} (Lf)^2 d\mu = \int_{\mathbb{R}^n} [(\nabla^2\psi)\nabla f \cdot \nabla f] d\mu + \int_{\mathbb{R}^n} \|\nabla^2 f\|_{HS}^2 d\mu \\
&\geq t \int_{\mathbb{R}^n} |\nabla f|^2 d\mu + \lambda \left[\int_{\mathbb{R}^n} |\nabla f|^2 d\mu - \left| \int_{\mathbb{R}^n} \nabla f d\mu \right|^2 \right] \\
&= (t+\lambda)\cdot\lambda - \lambda \left| \int_{\mathbb{R}^n} \nabla f d\mu \right|^2 . \qquad (3.10)
\end{aligned}$$

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

$$\left| \int_{\mathbb{R}^n} \nabla f d\mu \right|^2 \geq t. \qquad (3.11)$$

Using that the i^{th} coordinate of ∇f is $\nabla f \cdot \nabla x_i$ and integrating by parts we have

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

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

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

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

Observe that by testing the Poincaré inequality with linear functions, we obtain

$$\|\mathrm{Cov}(\mu)\|_{op} \leq C_P(\mu).$$

We thus deduce from Theorem 3.9 that

$$C_P(\mu) \leq \frac{1}{t}. \tag{3.12}$$

Inequality (3.12) is sometimes referred to as the log-concave Lichnerowicz inequality.

The Bochner formula states that in the log-concave case, for any $u \in C_c^\infty(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} (Lu)^2 d\mu = \int_{\mathbb{R}^n} [(\nabla^2\psi)\nabla u \cdot \nabla u] d\mu + \int_{\mathbb{R}^n} \|\nabla^2 u\|_{HS}^2 d\mu \geq \int_{\mathbb{R}^n} \|\nabla^2 u\|_{HS}^2 d\mu.$$

Let us dualize this inequality in order to obtain a Poincaré-type inequality. To this end, for $f \in L^2(\mu)$ we define the dual Sobolev norm

$$\|f\|_{H^{-1}(\mu)} = \sup\left\{\int_{\mathbb{R}^n} f u d\mu\,;\ \int_{\mathbb{R}^n} |\nabla u|^2 d\mu \leq 1\ u \in C_c^\infty(\mathbb{R}^n)\right\}.$$

This supremum can be finite only when $\int f d\mu = 0$.

Proposition 3.10. *(H^{-1}-inequality) Let μ be a regular, log-concave probability measure in $\mathbb{R}^n$. Then for $f \in L^2(\mu)$,*

$$Var_\mu(f) \leq \|\nabla f\|_{H^{-1}(\mu)}^2 = \sum_{i=1}^n \|\partial^i f\|_{H^{-1}(\mu)}^2.$$

Proof. We may assume that $\int f d\mu = 0$. By approximation, assume that $f = -Lu$ for $u \in C_c^\infty(\mathbb{R}^n)$. See [4] for the approximation argument. Then,

$$\int_{\mathbb{R}^n} f^2 d\mu = \int_{\mathbb{R}^n} [\nabla f \cdot \nabla u] d\mu \leq \|\nabla f\|_{H^{-1}(\mu)} \sqrt{\int_{\mathbb{R}^n} \|\nabla^2 u\|_{HS}^2 d\mu}$$
$$\leq \|\nabla f\|_{H^{-1}(\mu)} \sqrt{\int_{\mathbb{R}^n} (Lu)^2 d\mu}.$$

The proposition follows. □

The H^{-1}-norm has a geometric interpretation as infinitesimal *transport cost*, which may be roughly expressed by saying that when $\int f d\mu = 0$, as $\varepsilon \to 0$,

$$\|f\|_{H^{-1}(\mu)} \approx \frac{1}{\varepsilon} W_2(\mu, (1+\varepsilon f)\mu). \tag{3.13}$$

Let us explain (3.13). Let μ_1, μ_2 be Borel probability measures on $\mathbb{R}^n$. We say that a Borel probability measure γ on $\mathbb{R}^n \times \mathbb{R}^n$ is a *coupling* of μ_1 and μ_2 if

$$(\pi_i)_* \gamma = \mu_i \qquad\qquad (i = 1, 2),$$

where $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$ for $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. That is, the marginal of γ on the first coordinate is μ_1, and the marginal of γ on the second coordinate is μ_2. The L^2-Wasserstein distance between μ_1, μ_2 is defined as

$$W_2(\mu_1, \mu_2) = \inf_{\gamma} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \, d\gamma(x, y) \right)^{1/2}, \tag{3.14}$$

where the infimum runs over all couplings γ of μ_1 and μ_2. In probabilistic notation, we have

$$W_2(\mu_1, \mu_2) = \inf_{(X,Y)} \sqrt{\mathbb{E}|X - Y|^2}$$

where the infimum runs over all *possibly-dependent* random vectors $X, Y \in \mathbb{R}^n$ with X having law μ_1 and Y having law μ_2.

Proposition 3.11 ("bounding the H^{-1}-norm by transport cost"). *Let μ be a finite, compactly-supported measure on $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded, measurable function with*

$$\int f d\mu = 0.$$

For a sufficiently small $\varepsilon > 0$, let μ_ε be the measure whose density with respect to μ is the non-negative function $1 + \varepsilon f$. Then,

$$\|f\|_{H^{-1}(\mu)} \le \liminf_{\varepsilon \to 0^+} \frac{W_2(\mu, \mu_\varepsilon)}{\varepsilon}.$$

Proof. We need to prove that for any $u \in C_c^\infty(\mathbb{R}^n)$, function $u : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} f u d\mu \le \sqrt{\int_{\mathbb{R}^n} |\nabla u|^2 d\mu} \cdot \liminf_{\varepsilon \to 0^+} \frac{W_2(\mu, \mu_\varepsilon)}{\varepsilon}. \tag{3.15}$$

Fix such a test function $u \in C_c^\infty(\mathbb{R}^n)$. Then the second derivatives of u are bounded on $\mathbb{R}^n$. By Taylor's theorem, there exists a constant $R = R(u)$ with

$$u(y) - u(x) \le |\nabla u(x)| \cdot |x - y| + R|x - y|^2 \qquad \forall x, y \in \mathbb{R}^n. \tag{3.16}$$

We may assume that $\sup|f| > 0$ (otherwise, the theorem holds trivially), and let $\varepsilon > 0$ be smaller than $1/\sup|f|$. Then μ_ε is a non-negative measure on $\mathbb{R}^n$. Let γ be any coupling of μ and μ_ε. We see that

$$\int_{\mathbb{R}^n} fud\mu = \frac{1}{\varepsilon}\int_{\mathbb{R}^n} ud\left[\mu_\varepsilon - \mu\right] = \frac{1}{\varepsilon}\int_{\mathbb{R}^n\times\mathbb{R}^n} \left[u(y) - u(x)\right] d\gamma(x,y).$$

Write

$$W_2^\gamma(\mu,\mu_\varepsilon) = \sqrt{\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 d\gamma(x,y)}.$$

According to (3.16) and to the Cauchy-Schwarz inequality,

$$\begin{aligned}\int_{\mathbb{R}^n} hud\mu \quad &\le \quad \frac{1}{\varepsilon}\int_{\mathbb{R}^n\times\mathbb{R}^n} |\nabla u(x)|\cdot|x-y|d\gamma(x,y) + \frac{R}{\varepsilon}\int_{\mathbb{R}^n\times\mathbb{R}^n}|x-y|^2 d\gamma(x,y)\\ &\le \quad \frac{1}{\varepsilon}\sqrt{\int_{\mathbb{R}^n}|\nabla u(x)|^2 d\mu(x)}\cdot W_2^\gamma(\mu,\mu_\varepsilon) + \frac{R}{\varepsilon}W_2^\gamma(\mu,\mu_\varepsilon)^2.\end{aligned}$$

By taking the infimum over all couplings γ of μ and μ_ε, we obtain

$$\int_{\mathbb{R}^n} hud\mu \le \sqrt{\int_{\mathbb{R}^n}|\nabla u|^2 d\mu}\cdot\frac{W_2(\mu,\mu_\varepsilon)}{\varepsilon} + R\frac{W_2(\mu,\mu_\varepsilon)^2}{\varepsilon}, \tag{3.17}$$

with R depending only on u. We may assume that $\liminf_{\varepsilon\to 0^+} W_2(\mu,\mu_\varepsilon)/\varepsilon < \infty$; otherwise, there is nothing to prove. Consequently,

$$\liminf_{\varepsilon\to 0^+}\frac{W_2(\mu,\mu_\varepsilon)^2}{\varepsilon} = \liminf_{\varepsilon\to 0^+}\varepsilon\left(\frac{W_2(\mu,\mu_\varepsilon)}{\varepsilon}\right)^2 = 0.$$

Hence by letting ε tend to zero in (3.17), we deduce (3.15). The proof is complete. □

Exercises.

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

(2) Verify that the Poincaré constant of the standard Gaussian measure in $\mathbb{R}^n$ equals one.

(3) Let $t > 0$, and let X be a t-uniformly log-concave random vector in $\mathbb{R}^n$, Use the Prékopa-Leindler inequality, and show that for any subspace $E \subseteq \mathbb{R}^n$, also

$$Proj_E X$$

is a t-uniformly log-concave random vector.

(4) The Bochner formula also states that for any $u \in C_c^\infty(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} (Lu)^2 d\mu \geq \int_{\mathbb{R}^n} [(\nabla^2\psi)\nabla u \cdot \nabla u] d\mu.$$

Dualize this inequality in order to prove the *Brascamp-Lieb inequality*: For any C^1-smooth $f \in L^2(\mu)$,

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

Can you find equality cases, other than a constant function f?

(5) The Maurey-Pisier proof of Gaussian concentration.

   (a) Let X and Y be two independent, standard Gaussian random vectors in $\mathbb{R}^n$. For $\theta \in [0, \pi/2]$ set

   $$X_\theta = (\sin\theta)X + (\cos\theta)Y.$$

   Prove that $(X_\theta, \partial X_\theta/\partial\theta)$ coincides in distribution with (X, Y).

   (b) Let $F : \mathbb{R}^n \to \mathbb{R}$ be a locally-Lipschitz function and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Prove that

   $$\mathbb{E}\varphi(F(X) - F(Y)) = \mathbb{E}\varphi\left(\int_0^{\pi/2} \left\langle \nabla F(X_\theta), \frac{\partial X_\theta}{\partial\theta} \right\rangle d\theta\right).$$

   (c) Denote $E = \mathbb{E}F(X)$. Conclude that for any $\lambda > 0$,

   $$\mathbb{E}e^{\lambda(F(X)-E)} \leq \mathbb{E}e^{\lambda\pi\langle\nabla F(X),Y\rangle/2} = \mathbb{E}e^{\lambda^2\pi^2|\nabla F(X)|^2/8}.$$

(d) Conclude that if F is 1-Lipschitz, then for all $t > 0$,

$$\mathbb{P}\left(|F(X) - E| \geq t\right) \leq 2e^{-2t^2/\pi^2}.$$

(6) Let $\mu, \mu_1, \mu_2, \ldots$ be log-concave probability measures on $\mathbb{R}^n$. Assume that $\mu_N \longrightarrow \mu$ weakly, i.e., that for any continuous, compactly-supported function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\lim_{N\to\infty} \int_{\mathbb{R}^n} \varphi d\mu_N = \int_{\mathbb{R}^n} \varphi d\mu.$$

Suppose that μ_N is log-concave for all N. Prove that μ is log-concave.

(7) Complete the proof of Proposition 3.3

(8) Let $\rho : \mathbb{R}^n \to [0, \infty)$ be a log-concave probability density. Prove that there exist $A, B > 0$ such that for all $x \in \mathbb{R}^n$,

$$\rho(x) \leq Ae^{-B|x|}.$$

Lecture 4

The thin-shell bound under convexity assumption

Yesterday we discussed the thin-shell theorem, asserting that under the isotropic normalization, random vectors whose mass is concentrated in a *thin spherical shell* admit approximately Gaussian marginals. In this lecture we discuss the main ideas in the proof of the following thin shell bound:

Theorem 4.1. *Let X be an isotropic, log-concave random vector in $\mathbb{R}^n$. Then,*

$$\mathrm{Var}(|X|^2) = \mathbb{E}\left(|X|^2 - n\right)^2 \leq Cn, \tag{4.1}$$

where $C > 0$ is a universal constant.

By (4.1), for any isotropic, log-concave random vector X in $\mathbb{R}^n$,

$$\mathbb{E}(|X| - \sqrt{n})^2 \leq \frac{1}{n}\mathbb{E}(|X|^2 - n)^2 = \frac{1}{n}\mathrm{Var}(|X|^2) \leq C. \tag{4.2}$$

Hence most of the mass of the random vector X is located in a thin-spherical shell of radius $\sqrt{n}$ and width C.

Theorem 4.1 is tight, up to the value of the universal constant. Indeed, if X is a standard Gaussian random vector in $\mathbb{R}^n$ or if X is distributed uniformly in the cube $[-\sqrt{3}, \sqrt{3}]^n \subseteq \mathbb{R}^n$, then X is isotropic and log-concave with

$$\mathrm{Var}(|X|^2) = Cn,$$

where $C = 2$ in the Gaussian case and $C = 4/5$ in the case of the cube.

Remark 4.2. Theorem 4.1 and the Bourgain-Milman inequality imply an affirmative answer to *Bourgain's slicing problem*. Bourgain's slicing problem has played a highly influential role in the development of the theory of high-dimensional probability measures with convexity properties. An equivalent formulation, now established as a theorem, is the following *corrected form of the Busemann-Petty conjecture*: Let $n \geq 2$ and let $K, T \subseteq \mathbb{R}^n$ be centered convex bodies such that for any hyperplane $H \subseteq \mathbb{R}^n$ through the origin,

$$\mathrm{Vol}_{n-1}(K \cap H) \leq \mathrm{Vol}_{n-1}(T \cap H). \tag{4.3}$$

Then

$$\mathrm{Vol}_n(K) \leq C \cdot \mathrm{Vol}_n(T) \tag{4.4}$$

for a universal constant $C > 0$. For background on the slicing problem, see Ball [3] and Klartag and Milman [37]. For the resolution of Bourgain's problem in the affirmative, building on Guan's breakthrough [27], see Klartag and Lehec [39]. For a deduction of Bourgain's slicing theorem from Theorem 4.1 and the Bourgain-Milman inequality, see Eldan and Klartag [21].

We proceed with the main ideas of the proof of Theorem 4.1. Let μ be a probability measure on $\mathbb{R}^n$ with a log-concave density. Recall the $H^{-1}(\mu)$-norm, whose geometric meaning is understood through the *infinitesimal transport cost* bound

$$\|f\|_{H^{-1}(\mu)} \leq \liminf_{\varepsilon \to 0^+} \frac{W_2(\mu, (1+\varepsilon f)\mu)}{\varepsilon}, \tag{4.5}$$

where W_2 is the L^2-Wasserstein distance. The bound(4.5) is valid under minimal regularity assumptions on f, provided that $\int f d\mu = 0$. Recall the H^{-1}-inequality

$$\mathrm{Var}_\mu(f) \leq \sum_{i=1}^n \|\partial^i f\|^2_{H^{-1}(\mu)} \tag{4.6}$$

that holds for any smooth function $f \in L^2(\mu)$. In particular, by substituting $f(x) = |x|^2$ in (4.6) and noting that $\partial^i f = 2x_i$, Theorem 4.1 follows from the following:

Theorem 4.3. *Let μ be an isotropic, log-concave probability measure in $\mathbb{R}^n$. Then,*

$$\sum_{i=1}^n \|x_i\|^2_{H^{-1}(\mu)} \leq Cn,$$

where $C > 0$ is a universal constant.

Write p for the log-concave density of the probability measure μ in $\mathbb{R}^n$. For $y \in \mathbb{R}^n$, the corresponding *exponential tilt* of μ is the probability measure μ_y with density

$$p_y(x) = e^{x\cdot y - \Lambda(y)} p(x) \qquad\qquad (x \in \mathbb{R}^n), \tag{4.7}$$

where

$$\Lambda(y) = \log \int_{\mathbb{R}^n} e^{x\cdot y} d\mu(x)$$

is the logarithmic Laplace transform. Observe that for any $x \in \mathbb{R}^n$ and $i = 1, \ldots, n$, as $\varepsilon \to 0$,

$$p_{\varepsilon e_i}(x) = (1 + \varepsilon x_i)p(x) + o(\varepsilon).$$

It is an exercise to modify the proof of (4.5) and show that when μ is compactly-supported and $i = 1, \ldots, n$,

$$\|x_i\|_{H^{-1}(\mu)} \leq \limsup_{\varepsilon \to 0^+} \frac{W_2(\mu, \mu_{\varepsilon e_i})}{\varepsilon}. \tag{4.8}$$

Thus, in order to prove Theorem 4.3, it suffices to construct efficient *couplings* of exponential tilts of μ. The specific construction that we use for coupling tilts is related to the theory of non-linear filtering and to Eldan's stochastic localization from [20], which we now describe.

For simplicity, let us assume that the log-concave probability measure μ is compactly-supported. For $t \geq 0$ and $y \in \mathbb{R}^n$ we consider the t-log-Laplace transform

$$\Lambda_t(y) = \log \int_{\mathbb{R}^n} e^{y \cdot x - t|x|^2/2} p(x) dx,$$

and the *t-localized tilts* or *t-Gaussian needles*, which are the probability densities:

$$p_{t,y}(x) = e^{y \cdot x - t|x|^2/2 - \Lambda_t(y)} p(x) = \frac{p(x)\gamma_{1/t}(y/t - x)}{p * \gamma_{1/t}(y/t)}, \tag{4.9}$$

where

$$\gamma_s(x) = (2\pi s)^{-n/2} \exp(-|x|^2/(2s))$$

is the density of a centered Gaussian in $\mathbb{R}^n$ of covariance $s \cdot \mathrm{Id}$. The main advantage of the $p_{t,y}$ over the exponential tilt p_y is that $p_{t,y}$ is t-uniformly log-concave. In fact, almost everywhere in $\mathbb{R}^n$,

$$\nabla^2(-\log p_{t,y}) \geq t \cdot \mathrm{Id}. \tag{4.10}$$

Denote by $a_t(y)$ the barycenter of the probability density $p_{t,y}$, namely

$$a_t(y) = \nabla \Lambda_t(y) = \int_{\mathbb{R}^n} x p_{t,y}(x) dx \in \mathbb{R}^n. \tag{4.11}$$

It is an exercise to show that $a_t : \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz map, with a Lipschitz constant bounded uniformly in $t \in [0, +\infty)$. For reasons to be clarified soon, we are interested in the following integral equation:

Lemma 4.4. *For any continuous path $w = (w_t)_{t \geq 0}$ in $\mathbb{R}^n$ with $w_0 = 0$ and for any initial condition $\theta_0 \in \mathbb{R}^n$, there exists a unique solution $(\theta_t)_{t \geq 0}$ to the integral equation*

$$\theta_t = \theta_0 + w_t + \int_0^t a_s(\theta_s) ds, \qquad t \geq 0. \tag{4.12}$$

The solution $\theta_t = \theta_t(x)$ is continuous in $(t, x) \in [0, \infty) \times \mathbb{R}^n$ and is smooth in $x \in \mathbb{R}^n$ for any fixed $t \geq 0$. We denote

$$G_{t,w}(\theta_0) = \theta_t.$$

Thanks to the Lipschitz property of the map $a_t : \mathbb{R}^n \to \mathbb{R}^n$, Lemma 4.4 follows from standard Ordinary Differential Equations (ODE) theory; see [40], also for the standard fact that the map

$$G_{t,w} : \mathbb{R}^n \to \mathbb{R}^n \tag{4.13}$$

is a diffeomorphism. Consider a standard Brownian motion in $\mathbb{R}^n$

$$(W_t)_{t \geq 0},$$

with $W_0 = 0$. We will use the continuous Brownian path $W = (W_t)_{t\geq 0}$ in Lemma 4.4. Abbreviate

$$G_t(y) = G_{t,W}(y).$$

Proposition 4.5. *For any $y \in \mathbb{R}^n$, the stochastic process $(G_t(y))_{t\geq 0}$ has the same law as the process*

$$(y + tX_y + W_t)_{t\geq 0}$$

where X_y is a random vector with law μ_y which is independent of the Brownian motion $(W_t)_{t\geq 0}$.

Proposition 4.5 is part of the theory of non-linear filtering.

Corollary 4.6. *For any $y \in \mathbb{R}^n$, almost surely, the limit*

$$\lim_{t\to\infty} \frac{G_t(y)}{t} \tag{4.14}$$

exists, and has law μ_y.

Indeed, thanks to Proposition 4.5, Corollary 4.6 follows from the fact that

$$\lim_{t\to\infty} \frac{y + tX_y + W_t}{t} = X_y + \lim_{t\to\infty} \frac{W_t}{t} = X_y,$$

which is a random vector with law μ_y. Thus the limit in (4.14), usually denoted by $a_\infty(y)$, provides *simultaneous coupling* of all of the tilts $(\mu_y)_{y\in\mathbb{R}^n}$.

Proof of Proposition 4.5. Our proof requires some familiarity with stochastic processes.

Step 1. Observe that it suffices to prove the proposition for $y = 0$, since switching from X_0 to X_y amounts to replacing the function

$$a_s(\theta)$$

by

$$a_s(\theta + y).$$

We may thus assume that $y = 0$ and abbreviate $X = X_0$. For $t \geq 0$ define

$$\theta_t = tX + W_t. \tag{4.15}$$

The random vector $\theta_t/t = X + W_t/t$ is a noisy observation of X, which typically gets more and more accurate as t increases. Since W_t/t is a centered Gaussian random vector of covariance Id/t, the density of θ_t/t equals

$$p * \gamma_{1/t}.$$

Our goal is to prove that $(\theta_t)_{t\geq 0}$ coincides in law with $(G_{t,B}(0))_{t\geq 0}$ for a standard Brownian motion $(B_t))_{t\geq 0}$ in $\mathbb{R}^n$ with $B_0 = 0$.

Step 2. What is the conditional law of X given θ_t? It follows from (4.15) that the joint density of $(X, \theta_t/t)$ in $\mathbb{R}^n \times \mathbb{R}^n$ is

$$(x, z) \mapsto p(x)\gamma_{1/t}(z - x) \qquad (x, z) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{4.16}$$

Hence the density of the conditional law of X given θ_t/t is the normalized fiber of the joint density in (4.16), namely the probability density

$$\frac{p(x)\gamma_{1/t}(\theta_t/t - x)}{p * \gamma_{1/t}(\theta_t/t)} = p_{t,\theta_t}(x),$$

where we used (4.9) in the last passage. Thus p_{t,θ_t} is the density of the conditional law of X given the observation θ_t (or given the observation θ_t/t). In fact, p_{t,θ_t} is also the density of the conditional law of X given all past observations $(\theta_s)_{0\leq s\leq t}$. Indeed, it is an exercise to show that the time reversal

$$B_t = tW_{1/t} \qquad (t > 0)$$

is also a standard Brownian motion in $\mathbb{R}^n$, with $W_t = tB_{1/t}$. Thus,

$$\begin{aligned}
&\mathrm{Law}\,(X|(sX + W_s)_{0<s\leq t}) = \mathrm{Law}\left(X|(X + sW_{1/s})_{s\geq 1/t}\right) \\
&= \mathrm{Law}\left(X|(X + B_s)_{s\geq 1/t}\right) = \mathrm{Law}\left(X|X + B_{1/t} \text{ and } (B_s - B_{1/t})_{s>1/t}\right) \\
&= \mathrm{Law}\left(X|X + B_{1/t}\right) = \mathrm{Law}\,(X|tX + W_t)\,,
\end{aligned}$$

since $B_s - B_{1/t}$ is independent of X and $X + B_{1/t}$. Writing $\mathcal{F}_t$ for the σ-algebra generated by $(\theta_s)_{0\leq s\leq t}$, we conclude that

$$\mathbb{E}[X|\mathcal{F}_t] = \int_{\mathbb{R}^n} x p_{t,\theta_t}(x)dx = a_t(\theta_t). \tag{4.17}$$

Step 3. For $t \geq 0$ define $B_t \in \mathbb{R}^n$ via the equation

$$\theta_t = B_t + \int_0^t a_s(\theta_s)ds. \tag{4.18}$$

Thus, for $t \geq 0$,

$$\theta_t = G_{t,B}(0),$$

where $B = (B_t)_{t\geq 0}$. It remains to prove that the *innovation process* $(B_t)_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}^n$. From (4.18) we see that

$$B_t = W_t + tX - \int_0^t a_s(\theta_s)ds = W_t + \int_0^t (X - \mathbb{E}[X|\mathcal{F}_s])\, ds. \tag{4.19}$$

By (4.18) the random vector B_t is measurable with respect to $\mathcal{F}_t$. Consequently,

$$B_t = W_t + \int_0^t v_s ds \tag{4.20}$$

where

$$\mathbb{E}[v_t|(B_s)_{0\leq s\leq t}] = 0,$$

and where for $s < t$ the increment $W_t - W_s$ is a centered, Gaussian random vector of covariance $(t-s)\mathrm{Id}$ that is independent of $(B_r)_{0\leq r\leq s}$ and of $(v_r)_{0\leq r\leq s}$. We see from (4.20) that $(B_t)_{t\geq 0}$ is a martingale whose quadratic variation is that of a standard Brownian motion. Hence it is a Brownian motion, by Lévy's characterization. □

Thus far we have shown that for any $x, y \in \mathbb{R}^n$,

$$W_2(\mu_x, \mu_y) \leq \sqrt{\mathbb{E}\left|\lim_{t\to\infty}\frac{G_t(x)}{t} - \lim_{t\to\infty}\frac{G_t(y)}{t}\right|^2}, \tag{4.21}$$

because the first limit in (4.21) has law μ_x while the second has law μ_y. The next proposition refines (4.21) by allowing to stop the processes at a finite time.

Lemma 4.7. *For $x, y \in \mathbb{R}^n$ and $t > 0$,*

$$W_2(\mu_x, \mu_y) \leq \frac{1}{t}\cdot\sqrt{\mathbb{E}\,|G_t(x) - G_t(y)|^2}.$$

Proof. For $t \geq 0$ and $y \in \mathbb{R}^n$ we denote by $A_t(y)$ the covariance matrix of the probability density $p_{t,y}$, that is,

$$A_t(y) = \nabla^2\Lambda_t(y) = \int_{\mathbb{R}^n} x\otimes x\, p_{t,y}(x)dx - a_t(y)\otimes a_t(y) \in \mathbb{R}^{n\times n}. \tag{4.22}$$

Recall from (4.10) that $p_{t,y}$ is uniformly log-concave. Thus by the log-concave Lichnerowicz inequality,

$$A_t(y) = \nabla^2\Lambda_t(y) \leq \frac{1}{t}\cdot \mathrm{Id}.$$

This concavity property implies contraction properties of the time-dependent stochastic gradient ascent from Lemma 4.4. That is, for $y_1, y_2 \in \mathbb{R}^n$,

$$\begin{aligned}\langle a_t(y_1) - a_t(y_2), y_1 - y_2\rangle &= \langle \nabla\Lambda_t(y_1) - \nabla\Lambda_t(y_2), y_1 - y_2\rangle \\ &= \int_0^1 \left\langle \nabla^2\Lambda_t(sy_1 + (1-s)y_2)(y_1 - y_2), y_1 - y_2\right\rangle ds \leq \frac{1}{t}\cdot|y_1 - y_2|^2.\end{aligned} \tag{4.23}$$

By Lemma 4.4,

$$G_t(x) - G_t(y) = x - y + \int_0^t [a_s(G_s(x)) - a_s(G_s(y))]\, ds.$$

Hence by (4.23),

$$\begin{aligned}\frac{d}{dt}|G_t(x) - G_t(y)|^2 &= 2\langle a_t(G_t(x)) - a_t(G_t(y)), G_t(x) - G_t(y)\rangle \\ &\leq \frac{2}{t}|G_t(x) - G_t(y)|^2.\end{aligned}$$

Equivalently,

$$\frac{d}{dt}\frac{|G_t(x) - G_t(y)|^2}{t^2} \leq 0.$$

Hence

$$\frac{|G_t(x) - G_t(y)|^2}{t^2} \geq \limsup_{s\to\infty}\frac{|G_s(x) - G_s(y)|^2}{s^2} = \left|\lim_{s\to\infty}\frac{G_s(x) - G_s(y)}{s}\right|^2,$$

where the limit exists almost surely. The conclusion now follows from (4.21). □

Recall that $G_t : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth diffeomorphism. Denote

$$M_t = G_t'(0) \in \mathbb{R}^{n\times n},$$

i.e., $M_t v = \partial_v G_t(0)$ for any $v \in \mathbb{R}^n$. We write $|M_t|^2$ for the sum of the squares of the n^2 entries of the matrix M_t.

Corollary 4.8. *For any centered, compactly-supported, log-concave probability measure μ and $t > 0$,*

$$\sum_{i=1}^n \|x_i\|^2_{H^{-1}(\mu)} \leq \frac{1}{t^2}\cdot\mathbb{E}|M_t|^2.$$

Proof. It follows from Lemma 4.7 that

$$\limsup_{\varepsilon\to 0}\frac{W_2^2(\mu, \mu_{\varepsilon e_i})}{\varepsilon^2} \leq \frac{1}{t^2}\limsup_{\varepsilon\to 0}\mathbb{E}\left|\frac{G_t(0) - G_t(\varepsilon e_i)}{\varepsilon}\right|^2.$$

It is explained in [40] that the dominated convergence theorem allows us to replace expectation and limit, and obtain that

$$\limsup_{\varepsilon\to 0}\frac{W_2^2(\mu,\mu_{\varepsilon e_i})}{\varepsilon^2}\le\frac{1}{t^2}\mathbb{E}\left|\lim_{\varepsilon\to 0}\frac{G_t(0)-G_t(\varepsilon e_i)}{\varepsilon}\right|^2=\frac{1}{t^2}\cdot\mathbb{E}|G_t'(0)e_i|^2.$$

Thus, by (4.8),

$$\sum_{i=1}^n\|x_i\|_{H^{-1}(\mu)}^2\le\frac{1}{t^2}\sum_{i=1}^n\mathbb{E}|G_t'(0)e_i|^2=\frac{1}{t^2}\cdot\mathbb{E}|M_t|^2.$$

□

In order to prove Theorem 4.3 we will substitute $t=1$ in Corollary 4.8, and analyze the growth of the matrix-valued process $(M_s)_{0\le s\le 1}$. This analysis is quite technical, and it is described in the Appendix below.

Remark 4.9. Why does it make sense to use the above stochastic processes in order to bound the Wasserstein distance between exponential tilts of μ? After all, it is well-known that under mild regularity assumptions, the Wasserstein distance W_2 between two probability measures μ_1 and μ_2, as defined in (3.14), is realized by the *Brenier map*. That is, if p_i is the density of μ_i for $i=1,2$, then there exists an essentially unique map T that pushes forward μ_1 to μ_2 such that

$$W_2(\mu_1,\mu_2)=\sqrt{\int_{\mathbb{R}^n}|Tx-x|^2d\mu_1(x)}.$$

By "essentially unique" we mean that T is uniquely determined up to a set whose μ_1-measure is zero. See e.g. Villani [55] and references therein for background on the Brenier map. The Brenier map T is also the essentially unique map of the form $T=\nabla\Phi$ for a convex function $\Phi:\mathbb{R}^n\to\mathbb{R}$ that pushes forward μ_1 to μ_2. It satisfies the partial differential equation of Monge-Ampère type:

$$p_2(\nabla\Phi(x))\det\nabla^2\Phi(x)=p_1(x)\qquad(x\in\mathbb{R}^n).\tag{4.24}$$

Why don't we use this optimal Brenier map in order to bound the Wasserstein distance between exponential tilts of a log-concave measure μ? The reason is that analytically, the Brenier map is rather hard to analyze. In high dimensions, it is challenging to extract useful quantitative information from a partial differential equation such as (4.24); see [36] for an exception.

The coupling that we chose above, while non-optimal in the Wasserstein sense, is easier to analyze and it respects both the isotropy assumption and the log-concavity.

Second, it is optimal in a certain ways. In fact, let X and $(W_t)_{t\geq 0}$ be as in the proof of Proposition 4.5. Consider all stochastic processes of the form

$$\eta_t = W_t + \int_0^t u_s ds, \tag{4.25}$$

where the stochastic process $(u_t)_{t\geq 0}$ is assumed adapted to the filtration of the Brownian motion $(W_t)_{t\geq 0}$. For any fixed $T > 0$, among all processes $(\eta_t)_{t\geq 0}$ of the form (4.25) such that

$$\mathrm{Law}(\eta_T) = \mathrm{Law}(TX + W_T),$$

the process $(\theta_t)_{t\geq 0}$ with $\theta_t = W_t + \int_0^t a_s(\theta_s)ds$ minimizes the energy

$$\mathbb{E}\int_0^T |u_s|^2 ds.$$

This is explained in Lehec [43]. Thus, the process $(\theta_t/t)_{t>0}$ is the minimal "perturbation" of the rescaled Brownian motion that leads to the law of $X + W_t/t$, which tends to X as $t \longrightarrow \infty$. Moreover, the formulae involving this drift are particularly convenient for analyzing exponential tilts of the given measure μ. This provides some justification for the strategy of coupling exponential tilts using this process.

Remark 4.10. A considerable strengthening of Theorem 4.1 is the Kannan-Lovasz-Simonovits (KLS) conjecture [34]. In one of its formulations, the conjecture suggests that for any isotropic, log-concave random vector X in $\mathbb{R}^n$ and any locally-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}f^2(X) < \infty$,

$$Var f(X) \leq C\mathbb{E}|\nabla f(X)|^2, \tag{4.26}$$

where $C > 0$ is a universal constant. Theorem 4.1 establishes (4.26) in the particular case where $f(x) = |x|^2$, though the general case remains open. See [38] for information about this conjecture, and for a proof of a variant of (4.26) where C is replaced by $C \log n$.

Exercises.

(1) Modify the proof of (4.5) and prove (4.8).

(2) Let $(W_t)_{t\geq 0}$ be a standard Brownian motion with $W_0 = 0$. Set $B_t = tW_{1/t}$ for $t > 0$ and $B_0 = 0$. Prove that $(B_t)_{t\geq 0}$ is again a standard Brownian motion in $\mathbb{R}^n$.

(3) Let μ be an absolutely-continuous, compactly-supported probability measure with density p in $\mathbb{R}^n$. Consider the vector $a_t(y)$ $(t \geq 0, y \in \mathbb{R}^n)$ defined in

(4.11) above. Prove that $a_t : \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz map, with a Lipschitz constant bounded uniformly in $t \in [0, +\infty)$.

(4) Recall the proof of the Hadamard perturbation lemma and of the Hardy-Littlewood-Polya inequality.

Appendix

In order to prove Theorem 4.3, we should understand the matrix-valued process $(M_t)_{t\geq 0}$ of the derivative at zero of the random diffeomorphism G_t. Recall from (4.22) that we denote

$$A_t(y) = \nabla^2 \Lambda_t(y) = \mathrm{Cov}(p_{t,y})$$

and let us further abbreviate

$$A_t = A_t(G_t(0)).$$

The integral equation of Lemma 4.4 states that

$$G_t(y) = y + W_t + \int_0^t \nabla \Lambda_s(G_s(y))ds.$$

By differentiating with respect y (see [40] for justification) we see that

$$G_t'(0) = \mathrm{Id} + \int_0^t \nabla^2 \Lambda_s(G_s(0))G_s'(0)ds = \mathrm{Id} + \int_0^t A_s M_s ds.$$

Consequently, we have the *product integral equation*

$$\begin{cases} M_0 = \mathrm{Id} \\ \frac{d}{dt} M_t = A_t M_t \end{cases} \tag{4.27}$$

The following lemma is a non-probabilistic bound for the solution of the product integral equation. Denote the eigenvalues of A_t, repeated according to their multiplicity, by

$$\lambda_1(t) \geq \lambda_2(t) \geq \ldots \geq \lambda_n(t) > 0.$$

Lemma 4.11. *For any $t > 0$,*

$$|M_t|^2 \leq \sum_{i=1}^n \exp\left(2\int_0^t \lambda_i(s)ds\right). \tag{4.28}$$

It is straightforward to verify that for $n = 1$, equality holds in (4.28). Rather than proving Lemma 4.11 along the lines of [40], we will prove the lemma by using the Hardy-Littlewood-Polya inequality (see e.g. [50]). This inequality states that when $b_1 \geq b_2 \geq \ldots \geq b_m$ are real numbers and $c_1, \ldots, c_n \in \mathbb{R}$ are such that

$$\sum_{i=1}^k b_i \leq \sum_{i=1}^k c_i \qquad (k = 1, \ldots, n), \tag{4.29}$$

then for any convex, increasing function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\sum_{i=1}^{n} \varphi(b_i) \leq \sum_{i=1}^{n} \varphi(c_i). \tag{4.30}$$

Denote the singular values of M_t by

$$e^{b_1(t)} \geq \ldots \geq e^{b_n(t)}. \tag{4.31}$$

The numbers $e^{2b_1(t)}, \ldots, e^{2b_n(t)}$ are the eigenvalues of $M_t^* M_t$. These are absolutely-continuous functions of t. The proof of Lemma 4.11 relies on the following:

Lemma 4.12. *For $k = 1, \ldots, n$ and for almost any $t > 0$,*

$$\frac{d}{dt} \sum_{i=1}^{k} b_i(t) \leq \sum_{i=1}^{k} \lambda_i(t).$$

Proof. Fix $t > 0$ at which $b_1(t), \ldots, b_n(t)$ are differentiable, which happens almost everywhere. By an approximation argument it suffices to prove the lemma under the additional assumption that the inequalities in (4.31) are strict. Since A_t is a symmetric matrix, it follows from (4.27) that

$$\frac{d}{dt} M_t^* M_t = 2 M_t^* A_t M_t. \tag{4.32}$$

From the singular value decomposition of the matrix M_t, there exists orthonormal bases $u_1, \ldots, u_n \in \mathbb{R}^n$ and $v_1, \ldots, v_n \in \mathbb{R}^n$ such that

$$M_t u_i = e^{b_i(t)} v_i \qquad (i = 1, \ldots, n).$$

In particular $M_t^* M_t u_i = e^{2b_i(t)} u_i$. According to (4.32) and the Hadamard perturbation lemma,

$$\frac{d}{dt} e^{2b_i(t)} = 2 M_t^* A_t M_t u_i \cdot u_i \qquad (i = 1, \ldots, n).$$

Thus

$$2 e^{2b_i(t)} \frac{d}{dt} b_i(t) = 2 \langle A_t M_t u_i, M_t u_i \rangle = 2 e^{2b_i(t)} \langle A_t v_i, v_i \rangle.$$

In particular,

$$\frac{d}{dt} \sum_{i=1}^{k} b_i(t) = \sum_{i=1}^{k} \langle A_t v_i, v_i \rangle \leq \sum_{i=1}^{k} \lambda_i(t),$$

by the min-max characterization of the eigenvalues of the symmetric matrix A_t. $\square$

Proof of Lemma 4.11. Since $b_i(0) = 0$ for all i, we learn from Lemma 4.12 that for $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} b_i(t) \leq \sum_{i=1}^{k} \int_0^t \lambda_i(s)ds. \tag{4.33}$$

Denote $b_i = b_i(t)$ and $c_i = \int_0^t \lambda_i(s)ds$. Then $b_1 \geq \ldots \geq b_n$, while condition (4.29) holds true thanks to (4.33). Set $\varphi(t) = e^{2t}$, a convex increasing function. According to (4.30),

$$\sum_{i=1}^{n} e^{2b_i(t)} \leq \sum_{i=1}^{n} \exp\left(2 \int_0^t \lambda_i(s)ds\right).$$

Recalling that $e^{2b_1(t)}, \ldots, e^{2b_i(t)}$ are the eigenvalues of $M_t^* M_t$, the lemma follows. □

To summarize, thus far we obtained the following:

Corollary 4.13. *For any centered, compactly-supported, log-concave probability measure μ and $t > 0$,*

$$\mathrm{Var}_\mu(|x|^2) \leq \sum_{i=1}^{n} \|x_i\|^2_{H^{-1}(\mu)} \leq \frac{1}{t^2} \cdot \sum_{i=1}^{n} \mathbb{E} \exp\left(2 \int_0^t \lambda_i(s)ds\right),$$

where

$$\frac{1}{t} \geq \lambda_1(t) \geq \lambda_2(t) \geq \ldots \geq \lambda_n(t) > 0$$

are the eigenvalues of the covariance matrix A_t of the probability density

$$p_t = p_{t,G_t(0)}.$$

Let μ be an isotropic, log-concave probability measure in $\mathbb{R}^n$ with density p. It is an exercise to show that for proving the thin-shell theorem we may approximate μ and assume that p is continuous and compactly-supported.

Recall that for $t \geq 0$ and $y \in \mathbb{R}^n$ we consider the probability density

$$p_{t,y}(x) = e^{y \cdot x - t|x|^2/2 - \Lambda_t(y)} p(x) \qquad (x \in \mathbb{R}^n) \tag{4.34}$$

where

$$\Lambda_t(y) = \log \int_{\mathbb{R}^n} e^{y \cdot x - t|x|^2/2} p(x) \mathrm{d}x$$

is a normalizing factor. The barycenter and covariance of $p_{t,y}$ are given by

$$a_t(y) = \nabla \Lambda_t(y) = \int_{\mathbb{R}^n} x p_{t,y}(x) \mathrm{d}x \in \mathbb{R}^n$$

and

$$A_t(y) = \nabla^2 \Lambda_t(y) = \mathrm{Cov}(p_{t,y}) \in \mathbb{R}^{n \times n}.$$

We would also need the symmetric 3-tensor

$$\nabla^3 \Lambda_t(y) = \int_{\mathbb{R}^n} (x - a_t(y))^{\otimes 3} p_{t,y}(x) \mathrm{d}x \in \mathbb{R}^{n \times n \times n}.$$

Recall that $p_{t,y}$ is t-uniformly log-concave, i.e., $\nabla^2(-\log p_{t,y}) \geq t \cdot \mathrm{Id}$ for almost every $y \in \mathbb{R}^n$. One of our main proof ingredients is the following:

Lemma 4.14. *Let $t > 0$ and suppose that X is a centered, t-uniformly log-concave random vector in $\mathbb{R}^n$. Let $\lambda_1, \ldots, \lambda_n \geq 0$ be the eigenvalues of $\mathrm{Cov}(X)$ and let $u_1, \ldots, u_n \in \mathbb{R}^n$ be a corresponding orthonormal basis of eigenvectors. Abbreviate $X_i = \langle X, u_i \rangle$. Then for $1 \leq k \leq n$ and $s > 0$,*

$$\sum_{i,j=1}^{n} (\mathbb{E} X_i X_j X_k)^2 1_{\{\lambda_i \vee \lambda_j \leq s\}} \leq 4 t^{-1/2} s^{3/2} \lambda_k, \tag{4.35}$$

where $a \vee b = \max\{a, b\}$, i.e., in (4.35) we only sum over i, j with $\max\{\lambda_i, \lambda_j\} \leq s$.

Proof. Write $E \subseteq \mathbb{R}^n$ for the subspace spanned by the vectors u_i for which $\lambda_i \leq s$. Let $Proj_E$ be the orthogonal projection operator onto E in $\mathbb{R}^n$. Denote

$$Y = Proj_E X.$$

It follows from the Prékopa-Leindler inequality that Y is also t-uniformly log-concave, and

$$\|\mathrm{Cov}(Y)\|_{op} \leq s.$$

The improved log-concave Lichnerowicz inequality thus implies that the Poincaré constant of Y, denoted by $C_P(Y)$, satisfies

$$C_P(Y) \leq \sqrt{\frac{s}{t}}. \tag{4.36}$$

Set

$$H = \mathbb{E}\left[X_k Y \otimes Y\right] \in \mathbb{R}^{n \times n}.$$

By the definition of the subspace E,

$$\sum_{i,j=1}^{n} (\mathbb{E} X_i X_j X_k)^2 1_{\{\lambda_i \vee \lambda_j \leq s\}} = \mathrm{Tr}(H^2) \tag{4.37}$$

Moreover, by using (4.36) and the Poincaré inequality,

$$\begin{aligned} \mathrm{Var}(\langle HY, Y \rangle) &\leq C_P(Y) \cdot \mathbb{E}|2HY|^2 \leq 4 t^{-1/2} s^{1/2} \cdot \mathrm{Tr}(H^2 \mathrm{Cov}(Y)) \\ &\leq 4 t^{-1/2} s^{3/2} \cdot \mathrm{Tr} H^2. \end{aligned} \tag{4.38}$$

On the other hand, since $\mathbb{E}X_k = 0$, the Cauchy-Schwarz inequality shows that

$$\begin{aligned}\mathrm{Tr}(H^2) &= \mathbb{E}X_k\langle HY, Y\rangle \leq (\mathbb{E}X_k^2)^{1/2} \cdot (\mathrm{Var}\langle HY, Y\rangle)^{1/2} \\ &= \lambda_k^{1/2} \cdot \sqrt{\mathrm{Var}\langle HY, Y\rangle}. \end{aligned} \tag{4.39}$$

From (4.38) and (4.39),

$$\sqrt{\mathrm{Var}\langle HY, Y\rangle} \leq 4t^{-1/2}s^{3/2}\lambda_k^{1/2}. \tag{4.40}$$

The conclusion of the lemma follows from (4.37), (4.39) and (4.40). □

Let $(W_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^n$ with $W_0 = 0$. Consider the stochastic process $(\theta_t)_{t\geq 0}$ from the last lecture, for whose definition we offer two alternatives:

(1) The first option is to introduce a random vector X in $\mathbb{R}^n$ with law μ, independent of the Brownian motion $(W_t)_{t\geq 0}$, and set

$$\theta_t = tX + W_t.$$

(2) The second option is to uniquely define $(\theta_t)_{t\geq 0}$ via the integral equation

$$\theta_t = \int_0^t a_s(\theta_s)\mathrm{d}s.$$

The two options coincide in law, as we have seen last week. Write $\mathcal{F}_t$ for the σ-algebra generated by $(\theta_s)_{0\leq s\leq t}$. When we say that τ is a *stopping time* we mean that for any $t > 0$, the event $\{\tau \leq t\}$ is measurable with respect to $\mathcal{F}_t$. Denote

$$p_t = p_{t,\theta_t}, \qquad a_t = a_t(\theta_t), \qquad A_t = A_t(\theta_t), \qquad \Lambda_t = \Lambda_t(\theta_t)$$

and write

$$\frac{1}{t} \geq \lambda_1(t) \geq \lambda_2(t) \geq \ldots \geq \lambda_n(t) > 0 \tag{4.41}$$

for the eigenvalues of the covariance matrix A_t, repeated according to their multiplicity. Since μ is isotropic, at $t = 0$ we have $A_0 = \mathrm{Id}$ and hence

$$\lambda_1(0) = \lambda_2(0) = \ldots = \lambda_n(0) = 1.$$

For any k, the eigenvalue $\lambda_k(t)$ equals 1 at time $t = 0$, and it is smaller than 1 at any time $t > 1$. In the interval $(0, 1)$, however, the eigenvalue $\lambda_1(t)$ is typically very large, see the example in the exercise below. In view of Corollary 9.10 from last week, the missing ingredient in the proof of the thin-shell theorem along the lines of [40] is the following:

Proposition 4.15. *We have*

$$\sum_{k=1}^{n} \mathbb{E} \exp \left(2 \int_0^1 \lambda_k(t) \mathrm{d}t \right) \le Cn,$$

where $C > 0$ is a universal constant.

The proof of Proposition 4.15 relies on the following proposition, which is a straightforward variant of a recent breakthrough bound by Guan [27].

Proposition 4.16. *For any $t > 0$ and any stopping time τ,*

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{P} \left(\lambda_k(t \wedge \tau) \ge 3 \right) \le C e^{-1/t^{\alpha}},$$

where $a \wedge b = \min\{a, b\}$ and where $C, \alpha > 0$ are universal constants.

It is conceivable that $\alpha = 1$ in Proposition 4.16, see [28]. Proposition 4.16 tells us that while a single eigenvalue may explode at some time $t \in (0, 1)$, it is unlikely that many eigenvalues are simultaneously large.

Proof of Proposition 4.15 assuming Proposition 4.16. For $k = 1, \dots, n$ consider the stopping time

$$\tau_k = \inf \left\{ t > 0 \, ; \, \lambda_k(t) \ge 3 \right\}.$$

For any fixed $t > 0$ and $i = 1, \dots, k$, under the event $\tau_k \le t$ we have

$$\lambda_i(t \wedge \tau_k) \ge \lambda_k(t \wedge \tau_k) \ge 3.$$

Hence, for $i = 1, \dots, k$,

$$\mathbb{P} \left(\tau_k \le t \right) \le \mathbb{P} \left(\lambda_i(t \wedge \tau_k) \ge 3 \right).$$

By adding these k inequalities and using Proposition 4.16, for any $t > 0$,

$$\begin{aligned} \mathbb{P} \left(\tau_k \le t \right) &\le \frac{1}{k} \sum_{i=1}^{k} \mathbb{P} \left(\lambda_i(t \wedge \tau_k) \ge 3 \right) \le \frac{1}{k} \sum_{i=1}^{n} \mathbb{P} \left(\lambda_i(t \wedge \tau_k) \ge 3 \right) \\ &\le C \frac{n}{k} \exp(-1/t^{\alpha}). \end{aligned} \tag{4.42}$$

Recall that $\alpha > 0$ is a universal constant. It follows from (4.42) that

$$\mathbb{E} \tau_k^{-2} \le C \left(1 + \log \frac{n}{k} \right)^{2/\alpha}. \tag{4.43}$$

Indeed, in view of (4.42) inequality (4.43) clearly holds if $k \geq n/2$. For $k < n/2$ we obtain from (4.72) that for $s \geq 2^{2/\alpha}$,

$$\mathbb{P}\left(\frac{\tau_k^{-2}}{(\log \frac{n}{k})^{2/\alpha}} \geq s\right) \leq C\frac{n}{k}\exp(-s^{\alpha/2} \cdot \log\frac{n}{k}) = C\left(\frac{n}{k}\right)^{1-s^{\alpha/2}} \leq Ce^{-\tilde{c}s^{\alpha/2}}.$$

By integrating over $2^{2/\alpha} \leq s < \infty$ we obtain (4.43) . Consequently, since $\lambda_k(t) \leq 1/t$,

$$\int_0^1 \lambda_k(t)\mathrm{d}t \leq 3(\tau_k \wedge 1) + \int_{\tau_k \wedge 1}^1 \frac{\mathrm{d}t}{t} \leq 3 - \log(\tau_k \wedge 1). \tag{4.44}$$

Therefore, by (4.43) and (4.44),

$$\begin{aligned} \mathbb{E}\sum_{k=1}^n \exp\left(2\int_0^1 \lambda_k(t)\mathrm{d}t\right) &\leq e^6 \cdot \mathbb{E}\sum_{k=1}^n \mathbb{E}\left[\tau_k^{-2} \vee 1\right] \leq C\sum_{k=1}^n \mathbb{E}[\tau_k^{-2} + 1] \\ &\leq Cn \cdot \frac{1}{n}\sum_{k=1}^n \left(1 + \log\frac{n}{k}\right)^{2/\alpha} \leq \tilde{C}n, \end{aligned} \tag{4.45}$$

where the last passage follows from the fact that the function $(1 + \log(1/x))^{2/\alpha}$ is monotone and integrable in $[0, 1]$, and the Riemann sum in (4.45) may be bounded by the integral. □

The proof of Proposition 4.16 requires rather elaborate analysis of the time evolution of the eigenvalues of the covariance matrix A_t. Write

$$\xi_{ij}(t) = (\xi_{ij1}(t), \xi_{ij2}(t) \dots, \xi_{ijn}(t)) \in \mathbb{R}^n$$

where

$$\xi_{ijk}(t) = \int_{\mathbb{R}^n} \langle x - a_t, u_i\rangle \cdot \langle x - a_t, u_j\rangle \cdot \langle x - a_t, u_k\rangle p_t(x)\mathrm{d}x \in \mathbb{R},$$

with $u_1(t), \dots, u_n(t) \in \mathbb{R}^n$ being any orthonormal basis of eigenvectors of A_t corresponding to the eigenvalues $\lambda_1(t) \geq \dots \geq \lambda_n(t)$. Let us fix a stopping time τ.

Lemma 4.17. *For any smooth, increasing function $f : [0, \infty) \to \mathbb{R}$ and almost any $t > 0$,*

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^n f(\lambda_i(t \wedge \tau)) \leq \frac{1}{2}\sum_{i,j=1}^n \mathbb{E}\left[|\xi_{ij}(t)|^2 \frac{f'(\lambda_i(t)) - f'(\lambda_j(t))}{\lambda_i(t) - \lambda_j(t)} \cdot 1_{\{t<\tau\}}\right], \tag{4.46}$$

where we interpret the quotient by continuity as $f''(\lambda_i(t))$ when $\lambda_i(t) = \lambda_j(t)$. Moreover, the function that is differentiated on the left-hand side of (4.46) is absolutely continuous in $t \in [0, \infty)$.

The expression in the right-hand side of (4.46) is reminiscent of the Daleckii-Krein formula for the second derivative of matrix functions. For a function $f : \mathbb{R} \to \mathbb{R}$ and a symmetric matrix A whose spectral decomposition is

$$A = \sum_{i=1}^n \lambda_i u_i \otimes u_i$$

for numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and an orthonormal basis $u_1, \ldots, u_n \in \mathbb{R}^n$ we write

$$f(A) = \sum_{i=1}^n f(\lambda_i) u_i \otimes u_i.$$

The Daleckii-Krein formula states that for any two symmetric matrices $A, H \in \mathbb{R}^{n\times n}$, as $\varepsilon \to 0$,

$$\mathrm{Tr} f(A + \varepsilon H) = \mathrm{Tr} f(A) + \varepsilon \cdot \mathrm{Tr}[f'(A)H] + \frac{\varepsilon^2}{2} \cdot \mathrm{Tr}[(B \circ H)H] + o(\varepsilon^2)$$

where $\circ$ is the Schur product or Hadamard product (i.e., entry-wise product), and

$$B = \sum_{i,j=1}^n \frac{f'(\lambda_i) - f'(\lambda_j)}{\lambda_i - \lambda_j} u_i \otimes u_j.$$

For $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ we write $(\nabla^3 \Lambda_t) v \in \mathbb{R}^{n\times n}$ for the symmetric matrix whose i, j entry is

$$\left[(\nabla^3 \Lambda_t) v\right]_{ij} = \sum_{k=1}^n \Lambda_{t,ijk} v_k$$

where $\Lambda_t = (\Lambda_{t,ijk})_{i,j,k=1,\ldots,n}$. Lemma 4.17 follows from the following identity:

Lemma 4.18. *For any smooth function $f : [0, \infty) \to \mathbb{R}$ and almost any $t > 0$,*

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^n f(\lambda_i(t\wedge\tau)) = \frac{1}{2}\sum_{i,j=1}^n \mathbb{E}\left[|\xi_{ij}(t)|^2 \frac{f'(\lambda_i(t)) - f'(\lambda_j(t))}{\lambda_i(t) - \lambda_j(t)} \cdot 1_{\{t<\tau\}}\right]$$
$$- \mathbb{E}\left[\sum_{i=1}^n \lambda_i^2(t) f'(\lambda_i(t)) \cdot 1_{\{t<\tau\}}\right].$$

Moreover, the function that is differentiated is absolutely-continuous in $t \in [0, +\infty)$.

Proof. We will prove this lemma by using Itô calculus and the "first option" above for the definition of $(\theta_t)_{t\geq 0}$, i.e.,

$$\theta_t = tX + W_t.$$

Recall from last week that for some Brownian motion $(B_t)_{t\geq 0}$ we have

$$d\theta_t = dB_t + a_t dt \tag{4.47}$$

and that

$$p_t = p_{t,\theta_t}$$

is the conditional law of X given $(\theta_s)_{0\le s\le t}$. Recall that $\mathcal{F}_t$ is the σ-algebra generated by $(\theta_s)_{0\le s\le t}$. Hence, for any continuous test function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} \varphi p_t = \mathbb{E}\left[\varphi(X)|\mathcal{F}_t\right]. \tag{4.48}$$

The stochastic process on the left-hand side of (4.48) is a martingale, since it represents conditional expectations with respect to a non-decreasing family of σ-algebras. In fact, since p is compactly-supported and continuous, it follows that for any $x \in \mathbb{R}^n$,

$$(p_t(x))_{t\ge 0} \tag{4.49}$$

is a martingale as well. Recalling that

$$p_t(x) = e^{\theta_t\cdot x - t|x|^2/2 - \Lambda_t(\theta_t)} p(x)$$

we may apply the Itô formula based on (4.47) and obtain the evolution equation of the martingale (4.49), namely

$$dp_t(x) = \langle x - a_t, dB_t\rangle p_t(x). \tag{4.50}$$

It follows from (4.50) that

$$da_t = d\left[\int_{\mathbb{R}^n} x p_t(x)\mathrm{d}x\right] = \int_{\mathbb{R}^n} x\langle x - a_t, dB_t\rangle p_t(x)\mathrm{d}x = A_t dB_t.$$

Thus,

$$d(a_t\otimes a_t) = (A_t dB_t \otimes a_t + a_t \otimes A_t dB_t) + A_t^2 dt$$

and consequently,

$$dA_t = d\left[\int_{\mathbb{R}^n}(x\otimes x)p_t(x)\mathrm{d}x\right] - d[a_t\otimes a_t] = (\nabla^3\Lambda_t)dB_t - A_t^2 dt.$$

Hence, for any stopping time τ,

$$dA_{t\wedge\tau} = 1_{\{t<\tau\}}\cdot\left[(\nabla^3\Lambda_t)dB_t - A_t^2 dt\right].$$

Consequently,

$$d\mathrm{Tr} f(A_{t\wedge\tau}) = 1_{\{t<\tau\}}\cdot \mathrm{Tr}\left[f'(A_t)(\nabla^3\Lambda_t)dB_t - f'(A_t)A_t^2 dt + \frac{1}{2}D_t dt\right], \tag{4.51}$$

where the Itô term equals

$$D_t = \sum_{i,j=1}^n |\xi_{ij}(t)|^2 \frac{f'(\lambda_i(t)) - f'(\lambda_j(t))}{\lambda_i(t)-\lambda_j(t)},$$

thanks to the Daleckii-Krein formula. By taking expectation the dB_t term in (4.51) vanishes, completing the proof. $\square$

Since the measure μ is compactly-supported, there exists $R > 0$ depending on μ such that

$$|\xi_{ij}(t)| \le R \qquad \text{for all } i, j \text{ and } t \ge 0.$$

It is an instructive exercise to use Lemma 4.17 with $f(x) = e^{\beta x}$ in order to prove that for all $0 < t < c_\mu$,

$$\mathbb{P}(\lambda_1(t \wedge \tau) \ge 2) \le e^{-\tilde{c}_\mu / t} \tag{4.52}$$

for some constants $c_\mu, \tilde{c}_\mu > 0$ depending on the compactly-supported measure μ.

Our next goal is to use Lemma 4.17 and prove a bootstrap estimate for a certain class of functions considered by Guan [27], which generalizes the class of functions $f(t) = t^q$ $(q \ge 3)$ considered in Chen [16].

Lemma 4.19. *Let $D > 1, r \in [2,3], t > 0$ and let τ be a stopping time. Suppose that $f : [0,\infty) \longrightarrow [0,\infty)$ is a smooth, increasing function such that*

$$\begin{cases} f(x) = x^2, & \forall x \ge r \\ f''(x) \le D^2 f(x), & \forall x \ge 0 \end{cases} \tag{4.53}$$

Then, for almost any $t > 0$,

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^n f(\lambda_i(t\wedge\tau)) \le C\left(\frac{1}{t} + \frac{D^2}{\sqrt{t}}\right)\cdot \mathbb{E}\sum_{i=1}^n f(\lambda_i(t\wedge\tau)). \tag{4.54}$$

where $C > 0$ is a universal constant.

Proof. Abbreviate $\lambda_i = \lambda_i(t)$ and $\xi_{ij} = \xi_{ij}(t)$. Since f is positive, by Lemma 4.17 it suffices to prove that

$$\sum_{i,j=1}^n |\xi_{ij}|^2 \frac{f'(\lambda_i) - f'(\lambda_j)}{\lambda_i - \lambda_j} \le C\left(\frac{1}{t} + \frac{D^2}{\sqrt{t}}\right)\cdot \sum_{i=1}^n f(\lambda_i). \tag{4.55}$$

Since the probability density p_t is t-uniformly log-concave, Lemma 4.14 shows that for any $s > 0$ and $k = 1, \dots, n$,

$$\sum_{i,j=1}^n \xi_{ijk}^2 1_{\{\lambda_i \vee \lambda_j \le s\}} \le 4t^{-1/2} s^{3/2} \lambda_k. \tag{4.56}$$

The bound (4.55) follows from several applications of (4.56) as well as from the bound

$$\lambda_i \le 1/t$$

which was discussed in (4.41).

Step 1. Since ξ_{ijk} is symmetric in i, j and k, by using (4.56) with $s = \lambda_i$ we see that

$$\sum_{i,j=1}^{n} |\xi_{ij}|^2 1_{\{\lambda_i \ge r\}} = \sum_{i,j,k} \xi_{ijk}^2 1_{\{\lambda_i \ge r\}} \le 3 \sum_i 1_{\{\lambda_i \ge r\}} \sum_{j,k} \xi_{ijk}^2 1_{\{\lambda_j \vee \lambda_k \le \lambda_i\}} \tag{4.57}$$
$$\le \frac{12}{\sqrt{t}} \sum_i \lambda_i^{5/2} 1_{\{\lambda_i \ge r\}} \le \frac{12}{t} \sum_i \lambda_i^2 1_{\{\lambda_i \ge r\}} \le \frac{12}{t} \sum_{i=1}^{n} f(\lambda_i).$$

Step 2. Consider the contribution to the left-hand side of (4.55) of all i, j with

$$\min\{\lambda_i, \lambda_j\} \ge r. \tag{4.58}$$

Since $f'(x) = 2x$ when $x \ge r$, this contribution equals

$$\sum_{i,j=1}^{n} \frac{f'(\lambda_i) - f'(\lambda_j)}{\lambda_i - \lambda_j} |\xi_{ij}|^2 1_{\{\min(\lambda_i,\lambda_j) \ge r\}} = 2 \sum_{i,j} |\xi_{ij}|^2 1_{\{\min(\lambda_i,\lambda_j) \ge r\}}$$
$$\le 2 \sum_{i,j} |\xi_{ij}|^2 1_{\{\lambda_i \ge r\}} \le \frac{24}{t},$$

where we used (4.57) in the last passage.

Step 3. Consider the contribution to the left-hand side of (4.55) of all i, j with

$$\lambda_i \le r, \lambda_j \ge r+1 \qquad \text{or} \qquad \lambda_j \le r, \lambda_i \ge r+1. \tag{4.59}$$

This contribution equals

$$2 \sum_{i,j=1}^{n} \frac{f'(\lambda_i) - f'(\lambda_j)}{\lambda_i - \lambda_j} |\xi_{ij}|^2 1_{\{\lambda_i \ge r+1\}} 1_{\{\lambda_j \le r\}}$$
$$\le \sum_{i,j} \frac{4\lambda_i}{\lambda_i - \lambda_j} |\xi_{ij}|^2 1_{\{\lambda_i \ge r+1\}} 1_{\{\lambda_j \le r\}} \le 16 \sum_{i,j} |\xi_{ij}|^2 1_{\{\lambda_i \ge r\}} \le \frac{16 \cdot 12}{t}$$

Here we used that $f' \ge 0$ as well as the fact that $\lambda_i/(\lambda_i - \lambda_j) \le r + 1 \le 4$ when $\lambda_j \le r$ and $\lambda_i \ge r + 1$, and in the last passage we used (4.57).

Step 4. Let us show that

$$\sum_{i,j=1}^{n} |\xi_{ij}|^2 f(\lambda_i) 1_{\{\lambda_i \vee \lambda_j \le r+1\}} \le \frac{C}{\sqrt{t}} \sum_{k=1}^{n} f(\lambda_k). \tag{4.60}$$

Write $a \vee b \vee c = \max\{a, b, c\}$. By applying (4.56) with $s = r + 1$, and recalling that $r \leq 3$,

$$\sum_{i,j,k=1}^{n} f(\lambda_i)\xi_{ijk}^2 1_{\{\lambda_i \vee \lambda_j \vee \lambda_k \leq r+1\}} \tag{4.61}$$
$$\leq \frac{4}{\sqrt{t}} \sum_i f(\lambda_i) \cdot (r+1)^{3/2}\lambda_i \cdot 1_{\{\lambda_i \leq r+1\}} \leq \frac{4^{7/2}}{\sqrt{t}} \sum_{i=1}^{n} f(\lambda_i).$$

Next, we use that if $\lambda_i \leq r + 1$ then $f(\lambda_i) \leq f(r+1) = (r+1)^2 \leq 16$ while if $\lambda_i \geq r + 1$ then $f(\lambda_i) = \lambda_i^2$. We again apply (4.56) with $s = r + 1$ to obtain

$$\sum_{i,j,k=1}^{n} f(\lambda_i)\xi_{ijk}^2 1_{\{\lambda_i \vee \lambda_j \leq r+1 \leq \lambda_k\}} \leq \frac{C}{\sqrt{t}} \sum_k \lambda_k 1_{\{\lambda_k \geq r+1\}} \tag{4.62}$$
$$\leq \frac{C'}{\sqrt{t}} \sum_k \lambda_k^2 1_{\{\lambda_k \geq r+1\}} \leq \frac{C'}{\sqrt{t}} \sum_{k=1}^{n} f(\lambda_k).$$

By adding (4.61) and (4.62) we obtain (4.60).

Step 5. Consider the contribution to the left-hand side of (4.55) of all i, j with

$$\max\{\lambda_i, \lambda_j\} \leq r + 1. \tag{4.63}$$

By using (4.53) and the fact that f is non-negative and increasing, we see that this contribution is at most

$$2 \sum_{i,j=1}^{n} \frac{f'(\lambda_i) - f'(\lambda_j)}{\lambda_i - \lambda_j} |\xi_{ij}|^2 1_{\{\lambda_j \leq \lambda_i \leq r+1\}} \leq 4D^2 \sum_{i,j} f(\lambda_i)|\xi_{ij}|^2 1_{\{\lambda_j \leq \lambda_i \leq r+1\}}$$
$$\leq C \frac{D^2}{\sqrt{t}} \sum_{k=1}^{n} f(\lambda_k),$$

where we used (4.60) in the last passage.

The results of Step 2, Step 3 and Step 5 imply the desired bound (4.55). □

It is a calculus exercise to prove that for any $r \in [2, 3]$ and $D > 1$ there exists a smooth, increasing function $f : [0, \infty) \to (0, \infty)$ with

$$f(x) = \begin{cases} e^{D(x-r)} & x \leq r - \frac{1}{D} \\ x^2 & x \geq r \end{cases} \tag{4.64}$$

and $f''(x) \le D^2 f(x)$ for all $x \ge 0$. We denote this function f by $f_{r,D}$, and observe that it satisfies condition (4.53) of From Lemma 4.19. From the conclusion of the lemma we conclude that for any $D > 1, 2 \le r \le 3$ and a stopping time τ, if

$$0 < t \le D^{-4} \tag{4.65}$$

then $D^2/\sqrt{t} \le 1/t$ and hence for $f = f_{r,D}$

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^{n} f(\lambda_i(t \wedge \tau)) \le \frac{C}{t} \cdot \mathbb{E}\sum_{i=1}^{n} f(\lambda_i(t \wedge \tau)). \tag{4.66}$$

The function $f = f_{r,D}$ is slightly complicated, and we prefer to reformulate the growth condition (4.66) in terms of the much simpler function

$$g_r(x) = x^2 \cdot 1_{\{x \ge r\}}.$$

From (4.64),

$$g_r \le f_{r,D}. \tag{4.67}$$

In the other direction, we claim that for any $D > 1$ and $x \ge 0$, if

$$2 \le r + \frac{1}{D} \le \tilde{r} \le 3$$

then

$$f_{\tilde{r},D}(x) \le \frac{9}{4} g_r(x) + \exp(-D(\tilde{r} - r)). \tag{4.68}$$

Indeed, if $x \le r$ then by (4.64), since $r \le \tilde{r} - 1/D$,

$$f_{\tilde{r}}(x) \le f_{\tilde{r}}(r) = \exp(-D(\tilde{r} - r)),$$

and (4.68) holds true in this case. If $x \ge \tilde{r}$ then both $f_{\tilde{r}}(x)$ and $g_r(x)$ equal x^2, and (4.68) trivially holds. In the remaining case $r < x < \tilde{r}$ we have

$$f_{\tilde{r}}(x) \le f_{\tilde{r}}(\tilde{r}) = \tilde{r}^2 \le \left(\frac{\tilde{r}}{r}\right)^2 \cdot x^2 \le \frac{9}{4}x^2 = \frac{9}{4}g_r(x),$$

completing the proof of (4.68).

Proof of Proposition 4.16. We may assume that $t \le 2^{-8}$ as otherwise there is nothing to prove. We will set $t_0 = t$ and partition the interval $[0, t]$ into intervals

$$[t_1, t_0], [t_2, t_1], \dots, [t_{k+1}, t_k], \dots$$

For $k \ge 0$ we define

$$t_k = 2^{-8k} t, \qquad D_k = t_k^{-1/4}, \qquad r_k = 3 - \sum_{i=0}^{k-1} t_i^{1/8} \in [2, 3].$$

Since $t_k \le D_k^{-4}$ we may use the differential inequality (4.66) for all $s \in [t_{k+1}, t_k]$. By integrating this differential inequality over this interval, we obtain

$$\mathbb{E}\sum_{i=1}^{n} f_k(\lambda_i(t_k \wedge \tau)) \le \left(\frac{t_k}{t_{k+1}}\right)^C \mathbb{E}\sum_{i=1}^{n} f_k(\lambda_i(t_{k+1} \wedge \tau)), \tag{4.69}$$

where $f_k = f_{r_k, D_k}$. Set also $g_k = g_{r_k}$ and define

$$F_k = \mathbb{E}\sum_{i=1}^{n} f_k(\lambda_i(t_k \wedge \tau)) \qquad \text{and} \qquad G_k = \mathbb{E}\sum_{i=1}^{n} g_k(\lambda_i(t_k \wedge \tau)).$$

Note that $r_{k+1} + 1/D_k \le r_{k+1} + 1/\sqrt{D_k} = r_k$. From (4.69), as well as the two inequalities (4.67) and (4.68), we obtain for $k \ge 0$,

$$\begin{aligned} G_k \le F_k &\le \left(\frac{t_k}{t_{k+1}}\right)^C \mathbb{E}\sum_{i=1}^{n} f_{r_k, D_k}(\lambda_i(t_{k+1} \wedge \tau)) \\ &\le \left(\frac{t_k}{t_{k+1}}\right)^C \mathbb{E}\sum_{i=1}^{n} \left[\frac{9}{4} g_{k+1}(\lambda_i(t_{k+1} \wedge \tau)) + e^{-D_k(r_k - r_{k+1})}\right] \\ &= 2^{8C}\left(\frac{9}{4} G_{k+1} + n\exp(-t_k^{-1/8})\right) \le \bar{C}\left[G_{k+1} + n\exp(-2^k t^{-1/8})\right]. \end{aligned} \tag{4.70}$$

From this recursive inequality we obtain that for $k \ge 0$,

$$G_0 \le \bar{C}^k G_k + n \cdot \sum_{i=0}^{k-1} \bar{C}^{i+1} \exp(-2^i t^{-1/8}) \le \bar{C}^k G_k + \tilde{C} n \cdot e^{-t^{-1/8}}, \tag{4.71}$$

since the sum in (4.71) is at most

$$\sum_{i=0}^{k-1} \bar{C}^{i+1} \exp(-2^i t^{-1/8}) \le \sum_{i=0}^{\infty} \bar{C}^{i+1} \exp(-2^i - t^{-1/8}) = \tilde{C} \cdot e^{-t^{-1/8}}.$$

We next show that $\bar{C}^k G_k \longrightarrow 0$ as $k \to \infty$. To this end we use (4.52). Since μ is compactly-supported, for some $C_\mu > 0$ depending on μ and for a sufficiently large k,

$$G_k \le C_\mu \cdot \mathbb{P}\left(\lambda_1(t_k \wedge \tau) \ge 2\right) \le \tilde{C}_\mu e^{-c_\mu/t_k} = \tilde{C}_\mu e^{-c_\mu \cdot 2^{8k}/t}.$$

Hence indeed $\bar{C}^k G_k \longrightarrow 0$ as $k \to \infty$, and from (4.71),

$$\sum_{i=0}^{n} \mathbb{P}(\lambda_i(t \wedge \tau) \ge 3) \le G_0 \le \tilde{C} n \cdot e^{-t^{-1/8}}.$$

□

We end this lecture with an interpretation of our results in the context of the Prékopa-Leindler inequality. Recall that we write

$$\gamma_s(x) = (2\pi s)^{-n/2} \exp(-|x|^2/(2s))$$

for the density of a centered Gaussian random vector of covariance $s \cdot \mathrm{Id}$ in $\mathbb{R}^n$. Let p be an isotropic, log-concave density in $\mathbb{R}^n$ and for $t > 0$ set

$$q_t = p * \gamma_{1/t}.$$

By the Prékopa-Leindler inequality, the probability density q_t is log-concave, since it is a convolution of two log-concave probability measures. A straightforward computation based on (4.34) shows that

$$\nabla^2(-\log q_t)(x) = t^2 \left(\frac{\mathrm{Id}}{t} - \mathrm{Cov}(p_{t,tx}) \right) = t^2 \left(\frac{\mathrm{Id}}{t} - A_t(tx) \right).$$

Thus the log-concavity of q_t amounts to the inequality $A_t \leq \mathrm{Id}/t$, which was one of the starting points of our analysis today. By using the "Option 1" definition of θ_t, we see that for $t > 0$,

$$\int_{\mathbb{R}^n} \left| \frac{\mathrm{Id}}{t} - \frac{\nabla^2(-\log q_t)(x)}{t^2} \right|^2 q_t(x) dx = \mathbb{E}|A_t|^2 \leq Cn \tag{4.72}$$

where $|\cdot|$ is the Hilbert-Schmidt norm, and where the last inequality in (4.72) follows from Proposition 4.16. Thus, on a quantitative level, inequality (4.72) is a refinement of the Prékopa-Leindler inequality which amounts to the pointwise bound

$$0 \leq \nabla^2(-\log q_t) \leq t \cdot \mathrm{Id}.$$

Exercises.

(1) Why can we assume that μ is compactly-supported when proving the thin-shell theorem?

(2) prove that for any $D > 1$ and $r \in [2, 3]$ there exists a smooth, increasing function $f : [0, \infty) \longrightarrow [0, \infty)$ satisfying (4.53).

(3) Consider the isotropic, log-concave probability density

$$p(x_1, \ldots, x_n) = 2^n e^{-\sum_{i=1}^n 2|x_i|}.$$

(a) Prove that in this case, for any $t > 0$ the matrix A_t is diagonal and its diagonal entries are independent and identically-distributed. Write Z_t for the (1, 1)-entry of A_t, and explain that its law does not depend on n.

(b) Prove that the support of the random variable Z_t is *not* uniformly bounded for all $t \in (0, 1)$.

(c) Prove that if $x > 0$ is such that $\mathbb{P}(Z_t \geq x) \geq 1/n$, then $\mathbb{E}\|A_t\|_{op} \geq x/2$.

(d) Conclude that $\sup_{0<t<1} \mathbb{E}\|A_t\|_{op} \geq \alpha_n$, for some sequence $\alpha_n \longrightarrow \infty$.

(4) Assume that μ is isotropic, compactly-supported probability measure in $\mathbb{R}^n$.

(a) Use Lemma 4.17 and show that there exists $R = R_\mu > 0$ such that for a convex, smooth, increasing function $f : [0, \infty) \to \mathbb{R}$, and almost all $t > 0$,

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^{n} f(\lambda_i(t)) \leq R \sum_{i,j=1}^{n} \mathbb{E}f''(\lambda_i(t)).$$

(b) For $\beta > 0$ and $t > 0$ define

$$F_{\beta,t} = \frac{1}{\beta}\log \mathbb{E}\sum_{i=1}^{n} e^{\beta\lambda_i(t)}.$$

Prove that

$$F_{\beta,t} \leq tR\beta + \frac{\log n}{\beta} + 1.$$

(c) Write $p = \mathbb{P}(\lambda_1(t) \geq 2)$. Prove that for $\beta \geq 2\log n$,

$$\log p \leq tR\beta^2 - \frac{\beta}{2}.$$

Set $\beta = 1/(4tR)$ and conclude that for a sufficiently small $t > 0$,

$$\mathbb{P}(\lambda_1(t) \geq 2) \leq \exp(-c_\mu/t)$$

for some $c_\mu > 0$ depending on μ.

References

[1] Bakry, D., Gentil, I., Ledoux, M., *Analysis and Geometry of Markov Diffusion Operators.* Springer, Berlin, 2014.

[2] Ball, K., *Cube slicing in* $\mathbb{R}^n$. Proc. Amer. Math. Soc., Vol. 97, no. 3, (1986), 465–473.

[3] Ball, K., *The legacy of Jean Bourgain in geometric functional analysis.* Bull. Amer. Math. Soc., Vol. 58, no. 2, (2021), 205–223.

[4] Barthe, F., Klartag, B., *Spectral gaps, symmetries and log-concave perturbations.* Bull. Hellenic Math. Soc., Vol. 64, (2020), 1–31.

[5] Benyamini, Y., *Two-point symmetrization, the isoperimetric inequality on the sphere and some applications.* Texas functional analysis seminar (1983–1984), Longhorn Notes, University of Texas Press, (1984), 53–76.

[6] Berndtsson, B., *Complex integrals and Kuperberg's proof of the Bourgain-Milman theorem.* Adv. Math., Vol. 388, Paper No. 107927, (2021), 10 pp.

[7] Berwald, L., *Verallgemeinerung eines Mittelwertsatzes von J. Favard für positive konkave Funktionen.* Acta Math., Vol. 79, (1947), 17–37.

[8] Bobkov, S. G., *Isoperimetric and analytic inequalities for log-concave probability measures*. Ann. Probab., Vol. 27, no. 4, (1999), 1903–1921.

[9] Bobkov, S. G., *Remarks on the growth of* L^p*-norms of polynomials.* Geometric aspects of functional analysis, Israel seminar, Lecture Notes in Math., Vol. 1745, Springer, (2000), 27–35.

[10] Bobkov, S. G., Chistyakov, G., Götze, F., *Normal approximation for weighted sums under a second-order correlation condition.* Ann. Probab., Vol. 48, no. 3, (2020), 1202–1219.

[11] Bobkov, S. G., Chistyakov, G., Götze, F., *Concentration and Gaussian approximation for randomized sums*. Probability Theory and Stochastic Modelling, Vol. 104, Springer, 2023.

[12] Borell, C., *Complements of Lyapunov's inequality.* Math. Ann., Vol. 205, (1973), 323–331.

[13] Bourgain, J., *On the distribution of polynomials on high-dimensional convex sets.* Geometric aspects of functional analysis, Israel seminar, Lecture Notes in Math., Vol. 1469, Springer, (1991), 127–137.

[14] Bourgain, J., Milman, V. D., *New volume ratio properties for convex symmetric bodies in* $\mathbb{R}^n$. Invent. Math., Vol. 88, no. 2, (1987), 319–340.

[15] Carleson, L., *Selected problems on exceptional sets.* D. Van Nostrand Co., 1967.

[16] Chen, Y., *An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture.* Geom. Funct. Anal. (GAFA), Vol. 31, no. 1, (2021), 34–61.

[17] Diaconis, P., Freedman, D., *Asymptotics of graphical projection pursuit.* Ann. Statist., Vol. 12, no. 3, (1984), 793–815.

[18] Diaconis, P., Freedman, D., *A dozen de Finetti-style results in search of a theory.* Ann. Inst. H. Poincaré Probab. Statist., Vol. 23, no. 2, (1987), 397–423.

[19] Eitan, Y., *The centered convex body whose marginals have the heaviest tails.* Studia Math., Vol. 274, no. 3, (2024), 201–215.

[20] Eldan, R., *Thin shell implies spectral gap via a stochastic localization scheme*. Geom. Funct. Anal. (GAFA), Vol. 23, (2013), 532–569.

[21] Eldan, R., Klartag, B., *Approximately Gaussian marginals and the hyperplane conjecture.* In Concentration, functional inequalities and isoperimetry, Vol. 545 of Contemp. Math., Amer. Math. Soc., (2011), 55–68.

[22] Feller, W., *An introduction to probability theory and its applications.* Vol. II. Second edition. John Wiley & Sons, 1971.

[23] Figiel, T., Lindenstrauss, J., Milman, V. D., *The dimension of almost spherical sections of convex bodies*. Acta Math., Vol. 139, no. 1-2, (1977), 53–94.

[24] Fradelizi, M., *Hyperplane sections of convex bodies in isotropic position.* Beiträge Algebra Geom., Vol. 40, no. 1, (1999), 163–183.

[25] Gromov, M., *Paul Lévy's isoperimetric inequality.* Appendix C in Gromov's book *Metric structures for Riemannian and non-Riemannian spaces*, Birkhäuser, 1999.

[26] Gromov, M., Milman, V. D., *Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces.* Compositio Math., Vol. 62, no. 3, (1987), 263–282.

[27] Guan, Q., *A note on Bourgain's slicing problem.* Preprint, arXiv:2412.09075

[28] Guan, Q., *On tail probability of the covariance matrix in Eldan's stochastic localization*. arXiv:2508.14943

[29] Hadwiger, H., Ohmann, D., *Brunn-Minkowskischer Satz und Isoperimetrie.* Math. Z., Vol. 66, (1956), 1–8.

[30] Hensley, D., *Slicing the cube in* $\mathbb{R}^n$ *and Probability (bounds for the measure of a central cube slice in* $\mathbb{R}^n$ *by probability methods).* Proc. Amer. Math. Soc., Vol. 73, no. 1, (1979), 95–100.

[31] Hensley, D., *Slicing convex bodies—bounds for slice area in terms of the body's covariance.* Proc. Amer. Math. Soc., Vol. 79, no. 4, (1980), 619–625.

[32] Henstock, R., Macbeath, A. M., *On the measure of sum-sets. I. The theorems of Brunn, Minkowski, and Lusternik.* Proc. London Math. Soc. (3), Vol. 3 (1953), 182–194.

[33] Iriyeh, H., Shibata, M., *Symmetric Mahler's conjecture for the volume product in the 3-dimensional case.* Duke Math. J., Vol. 169, no. 6, (2020), 1077–1134.

[34] Kannan, R., Lovász, L., Simonovits, M., *Isoperimetric problems for convex bodies and a localization lemma*. Discrete Comput. Geom., Vol. 13, no. 3-4, (1995), 541–559.

[35] Klartag, B., *Marginals of geometric inequalities.* Geometric aspects of functional analysis, Israel seminar, Lecture Notes in Math., Vol. 1910, Springer, (2007), 133–166.

[36] Klartag, B., *Logarithmically-concave moment measures I.* Geometric aspects of functional analysis, Israel seminar, Lecture Notes in Math., Vol. 2116, (2014), 231–260.

[37] Klartag, B., Milman, V., *The slicing problem by Bourgain*. Analysis at Large, Dedicated to the Life and Work of Jean Bourgain, edited by A. Avila, M. Rassias and Y. Sinai, Springer, (2022), 203–232.

[38] Klartag, B., *Logarithmic bounds for isoperimetry and slices of convex sets.* Ars Inveniendi Analytica, Paper No. 4, (2023), 17pp.

[39] Klartag, B., Lehec, J., *Affirmative resolution of Bourgain's slicing problem using Guan's bound*. Geom. Funct. Anal. (GAFA), Vol. 35, (2025), 1147–1168.

[40] Klartag, B., Lehec, J., *Thin-shell bounds via parallel coupling*. Preprint, arXiv:2507.15495

[41] Koldobsky, A., *Fourier analysis in convex geometry*. American Mathematical Society, 2005.

[42] Kuperberg, G., *From the Mahler Conjecture to Gauss Linking Integrals.* Geom. Funct. Anal. (GAFA), Vol. 18, (2008), 870–892.

[43] Lehec, J., *Representation formula for the entropy and functional inequalities.* Ann. Inst. Henri Poincaré Probab. Stat., Vol. 49, no. 3, (2013), 885–899.

[44] Mahler, K., *Ein Minimalproblem fur konvexe Polygone.* Mathematica Zutphen B7, (1939), 118–127.

[45] Mahler, K., *Ein Übertragungsprinzip für konvexe körper*. Časopis Pest Mat. Fys., Vol. 68, (1939), 93–102.

[46] McMullen, P., *Volumes of projections of unit cubes.* Bull. London Math. Soc., Vol. 16, no. 3, (1984), 278–280.

[47] Meyer, M., Pajor, A., *On the Blaschke-Santaló inequality*. Arch. Math., Vol. 55, (1990), 82–93.

[48] Müller, C., *Spherical Harmonics.* Lecture Notes in Math., Vol. 17, Springer, 1966.

[49] Nazarov, F., Podkorytov, A. N., *Ball, Haagerup, and distribution functions.* In Complex analysis, operators, and related topics, Oper. Theory Adv. Appl., Vol. 113, Birkhäuser, (2000), 247–267.

[50] Polya, G., *Remark on Weyl's note "Inequalities between the two kinds of eigenvalues of a linear transformation."* Proc. Nat. Acad. Sci. U. S. A., Vol. 36, (1950), 49–51.

[51] Santaló, L. A., *Un invariante afin para los cuerpos convexos del espacio den dimensiones.* Portugal. Math., Vol. 8, (1949), 155–161.

[52] Schmidt, E., *Der Brunn-Minkowskische Satz und sein Spiegeltheorem sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und hyperbolischen Geometrie.* Math. Ann., Vol. 120, (1947), 307–422.

[53] Sudakov, V. N., *Typical distributions of linear functionals in finite-dimensional spaces of high-dimension.* (Russian) Dokl. Akad. Nauk. SSSR, Vol. 243, no. 6, (1978), 1402–1405. English translation in Soviet Math. Dokl., Vol. 19, (1978), 1578–1582.

[54] Vaaler, J. D., *A geometric inequality with applications to linear forms.* Pacific J. Math., Vol. 83, no. 2, (1979), 543–553.

[55] Villani, C., *Topics in optimal transportation.* American Mathematical Society, 2003.

[56] Vershynin, R., *High-dimensional probability.* Cambridge University Press, 2018.