Isoperimetric inequalities in high-dimensional convex sets
Boaz Klartag, ETH Zurich 2025

Lecture 3: The thin shell theorem and the Johnson-Lindenstrauss lemma

In this lecture and the next, we discuss applications of the spherical concentration of measure phenomenon in high-dimensions and prove results that (arguably) shed light on the geometry of high-dimensional Euclidean space. We begin with the following corollary of Lévy's isoperimetric inequality:

Theorem 3.1 ("spherical concentration of Lipschitz functions")**.** *Let $f : S^{n-1} \to \mathbb{R}$ be an L-Lipschitz function, i.e., $|f(x) - f(y)| \leq L|x - y|$ 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 $t > 0$,

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

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

Theorem 3.1 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 3.1. We may assume that $L = 1$ (otherwise, replace f by f/L). Abbreviate $\{f \leq t\} = \{x \in S^{n-1} \, ; \, f(x) \leq t\}$. Let $M \geq 0$ be a median of the function f, i.e.,

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

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

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

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

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

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

$$\sigma_{n-1}(\{f \geq M - t\}) \geq 1 - Ce^{-ct^2 n}. \tag{3}$$

From (2) and (3), for any $t > 0$.

$$\sigma_{n-1}(\{|f - M| \geq t\}) \leq Ce^{-ct^2 n}. \tag{4}$$

The expectation of f is rather close to the median. In fact, by (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 $t > 0$,

$$\sigma_{n-1}(\{|f - E| \geq t\}) \leq Ce^{-ct^2 n}. \tag{5}$$

Indeed, if $t \leq 1/\sqrt{n}$ then the right-hand side of (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 (5) follows from (4). $\square$

As we see from the proof of Theorem 3.1, we may replace the expectation E in (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 3.2. 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

$$\int_{S^{n-1}} f^2 d\sigma_{n-1} \leq \frac{1}{n-1} \int_{S^{n-1}} |\nabla f|^2 d\sigma_{n-1}. \tag{6}$$

Equality holds in (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 [8]. There are also L^p-versions of the Poincaré inequality (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 (6) implies sub-exponential concentration of Lipschitz functions, which is considerably weaker than the sub-Gaussian concentration of Theorem 3.1.

3.1 The thin shell theorem

Our first application of the spherical concentration of measure phenomenon is a version of the "thin-shell theorem" of Sudakov [9] and Diaconis–Freedman [4]. This theorem offers additional insight into why the Gaussian distribution arises in the central limit theorem.

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

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

or in short if

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

where $x \otimes x = (x_i x_j)_{i,j=1,\ldots,n} \in \mathbb{R}^{n \times n}$ 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 3.3 (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 $\varepsilon \geq 0$,*

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \leq \varepsilon^2. \tag{7}$$

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 $t \in \mathbb{R}$,

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

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

The exponents $1/2$ and $1/8$ on the right-hand side of (8) are non-optimal. Bobkov, Chistyakov and Götze [2, 3] used the Fourier transform as well as other techniques, and essentially obtained $C\varepsilon^2 \log n$ on the right-hand side of (8), with a slightly different definition of ε, and with a slightly different probabilistic estimate on θ.

What is the meaning of condition (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 (7) implies that the bulk of the mass of X is concentrated in a *thin spherical shell.*

Theorem 3.3 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 3.3 one can see that the thin-shell condition (7) is also necessary for the Gaussian approximation phenomenon of the majority of the marginals.

Examples.

1. Consider the case where $X = (X_1, \ldots, X_n)$ and $X_1, \ldots, X_n$ are independent random variables with, say, $\mathbb{E}X_i^2 = 1$ and $\mathbb{E}X_i^4 \leq 100$ for all i. The thin-shell condition (7) holds true with a rather small ε. Indeed, we may compute that

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

   Thus the standard deviation of $|X|/\sqrt{n}$ is at most $10/\sqrt{n}$, and (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 3.3 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 3.3, most of the marginals of X are approximately Gaussian.

3. In the second part of this course, we will (hopefully) discuss a recent proof that the uniform distribution on any convex set in $\mathbb{R}^n$, when isotropic, satisfies the requirements of Theorem 3.3 with $\varepsilon = C/\sqrt{n}$, see Klartag and Lehec [6].

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 3.3 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

$$S^{n-1} \ni \theta \mapsto \mathbb{P}\left(X \cdot \theta \leq t\right) \tag{9}$$

for $t \in \mathbb{R}$. The function in (9) is not necessarily continuous in general, but as we will see it admits good Lipschitz approximations. We begin the proof of Theorem 3.3 with the following:

Lemma 3.4. *Let X and ε be as in Theorem 3.3, 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 $\sigma_{n-1}(\Theta) \geq 1 - C\exp(-c\sqrt{n})$ such that for any $\theta \in \Theta$,*

$$\left|\mathbb{E} f(X \cdot \theta) \;-\; \mathbb{E} f(\sqrt{n} Y_1)\right| \leq CL\left(\frac{1}{n^{1/4}} + \varepsilon\right). \tag{10}$$

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 (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 $\theta_1, \theta_2 \in S^{n-1}$,

$$|F(\theta_1) - F(\theta_2)| \leq \mathbb{E}\,|f(X \cdot \theta_1) - f(X \cdot \theta_2)| \leq L\mathbb{E}\,|X \cdot (\theta_1 - \theta_2)|$$
$$\leq L\sqrt{\mathbb{E}|X \cdot (\theta_1 - \theta_2)|^2} = L|\theta_1 - \theta_2|,$$

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 3.1 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

$$\forall \theta \in \Theta, \qquad \left| F(\theta) - \int_{S^{n-1}} F d\sigma_{n-1} \right| \leq \frac{L}{n^{1/4}}. \tag{11}$$

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

$$\langle X/|X|, Y \rangle \qquad \text{and} \qquad Y_1 \tag{12}$$

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 (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,

$$\int_{S^{n-1}} F d\sigma_{n-1} = \mathbb{E}F(Y) = \mathbb{E}f(X \cdot Y) = \mathbb{E}f(|X|Y_1). \tag{13}$$

Our main assumption (7) implies that the random variable $|X|$ is typically very close to $\sqrt{n}$. Thus,

$$\left|\mathbb{E}f(|X|Y_1) - \mathbb{E}f(\sqrt{n}Y_1)\right| \leq L \cdot \mathbb{E}\left|\left(|X| - \sqrt{n}\right) Y_1\right|$$
$$\leq L\sqrt{\mathbb{E}nY_1^2} \cdot \sqrt{\mathbb{E}\left(|X|/\sqrt{n} - 1\right)^2} \leq L\varepsilon,$$

as $\mathbb{E}Y_1^2 = 1/n$. Combining the last inequality with (11) and (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 $t \in \mathbb{R}$

$$\left|\mathbb{P}\left(\sqrt{n}Y_1 \leq t\right) - \mathbb{P}\left(Z \leq t\right)\right| \leq \frac{C}{n}. \tag{14}$$

Proof of Theorem 3.3. 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

$$\mathbb{P}(X \cdot \theta \leq t) \leq \mathbb{E}I_t(X \cdot \theta) \leq \mathbb{P}(X \cdot \theta \leq t + \delta). \tag{15}$$

From Lemma 3.4, for any $t \in \mathbb{R}$ there exists $\mathcal{A}_t \subseteq S^{n-1}$ with

$$\sigma_{n-1}(\mathcal{A}_t) \geq 1 - Ce^{-c\sqrt{n}} \tag{16}$$

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

$$\left|\mathbb{E}I_t(X \cdot \theta) - \mathbb{E}I_t(\sqrt{n}Y_1)\right| \leq C \cdot \frac{1}{\delta} \cdot \left(n^{-1/4} + \varepsilon\right) \leq \tilde{C}\sqrt{\delta}. \tag{17}$$

Our goal is to leverage (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 $t \in \mathbb{R}$,

$$|\mathbb{P}(X \cdot \theta \leq t) - \mathbb{P}(Z \leq t)| \leq C\sqrt{\delta}. \tag{18}$$

Step 1. We would like to replace $\sqrt{n}Y_1$ in (17) by Z. By the definition of I_t and by (14),

$$\begin{aligned} \mathbb{P}(Z \leq t) - C/n \leq \mathbb{P}\left(\sqrt{n}Y_1 \leq t\right) &\leq \mathbb{E}I_t(\sqrt{n}Y_1) \\ &\leq \mathbb{P}(\sqrt{n}Y_1 \leq t + \delta) \leq \mathbb{P}(Z \leq t + \delta) + \tilde{C}/n. \end{aligned} \tag{19}$$

Moreover,

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

Thus by (19),

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

Consequently, from (17), for any $\theta \in \mathcal{A}_t$,

$$|\mathbb{E}I_t(X \cdot \theta) \, - \, \mathbb{P}(Z \le t)| \le C\sqrt{\delta}. \tag{20}$$

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

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

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

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

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 (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 (18). Pick $\theta \in \mathcal{A}$ and $t \in \mathbb{R}$. There exists $j = 0, \ldots, k$ such that $t_j \le t \le t_{j+1}$. Thus, by (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 (18). For the other half, let $i = 0, \ldots, k$ 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 (18). ☐

3.2 The Johnson-Lindenstrauss lemma on dimension reduction

We move on to our second application of spherical concentration of measure. In their work [5] on Lipschitz extensions of maps from metric spaces to a Euclidean space, Johnson and Lindenstrauss proved the following lemma:

Lemma 3.5 ("The JL lemma on dimension reduction")**.** *Suppose that $X \subseteq \mathbb{R}^n$ is a finite set and let $0 < \varepsilon < 1/2$. Let k be an integer such that*

$$k \geq C\varepsilon^{-2} \cdot \log \#(X).$$

Then there exists a map $f : X \to \mathbb{R}^k$ such that for all $x, y \in X$,

$$\frac{1}{1+\varepsilon}|x-y| \leq |f(x) - f(y)| \leq (1+\varepsilon)|x-y|.$$

In fact, the map f is the restriction to X of a linear map from $\mathbb{R}^n$ to $\mathbb{R}^k$.

We say that the map f from Lemma 3.5 is a *metric embedding with distortion at most* $1+\varepsilon$, since it distorts distances by a *multiplicative* factor of at most $1+\varepsilon$. Thus any subset of N points in Euclidean space in any dimension may be embedded in dimension $C \log N$ while preserving all distances up to a factor of at most 1.01.

Example. Consider the $n+1$ vertices of a regular simplex in $\mathbb{R}^n$ with edgelength 1. That is, the points $x_0, \dots, x_n \in \mathbb{R}^n$ satisfy

$$|x_i - x_j| = \begin{cases} 1 & i \neq j \\ 0 & i = j \end{cases}$$

An exact metric embedding of this metric space $X = \{x_0, \dots, x_n\}$ is possible only in dimension n and above. However, an embedding with distortition of 1.01 is possible already in dimension $C \log n$. Putting it differently, in $\mathbb{R}^k$ and even in the unit sphere S^{k-1}, there is a collection of e^{ck} points where all distances between pairs of points lie in the interval

$$[0.99, 1.01].$$

This finite subset of S^{k-1} is an "approximate simplex" with exponentially many points.

In order to prove the JL lemma, we need to introduce the notion of a random k-dimensional subspace of $\mathbb{R}^n$. The collection of all k-dimensional subspaces of $\mathbb{R}^n$ is referred to as the Grassmannian $G_{n,k}$. The Grassmannian $G_{n,k}$ is a smooth manifold of dimension $k(n-(k+1)/2)$. The group $O(n)$ of all orthogonal transformations on $\mathbb{R}^n$ acts transitively on $G_{n,k}$, meaning that for any two given subspaces $E_1, E_2 \in G_{n,k}$, there is an element $U \in O(n)$ such that

$$U(E_1) = E_2.$$

In other words, the Grassmannian $G_{n,k}$ is a *homogenous space* of the orthogonal group. Haar's theorem states that since the compact group $O(n)$ acts transitively on $G_{n,k}$, there exists a unique rotationally-invariant probability measure on $G_{n,k}$. Moreover, there is a unique (Haar) rotationally-invariant probability measure also on $O(n)$, referred to as the uniform probability measure on $O(n)$.

Whenever we say that $E \in G_{n,k}$ is a random subspace, distributed uniformly in $G_{n,k}$, we refer to this Haar probability measure on $G_{n,k}$. Observe that for any fixed rotation $U \in O(n)$, the random subspace $U(E)$ is also uniformly distributed in the Grassmannian $G_{n,k}$. Similarly, if $U \in O(n)$ is a random orthogonal transformation, uniformly distributed on $O(n)$, then also $U^{-1} = U^*$ is uniformly distributed on $O(n)$, as well as UU_0 and U_0U for any fixed matrix $U_0 \in O(n)$. The uniqueness of the Haar measure has the following consequences:

Claim 3.6.

1. *If $X_1, \ldots, X_k \in S^{n-1}$ are independent random vectors, each distributed uniformly in S^{n-1}, then*

$$E = sp\{X_1, \ldots, X_k\}$$

   is almost surely a k-dimensional subspace, distributed uniformly in $G_{n,k}$.

2. *If $E_0 \in G_{n,k}$ is a fixed subspace, and $U \in O(n)$ is a random, uniformly-distributed orthogonal transformation, then*

$$U(E_0)$$

   is distributed uniformly in $G_{n,k}$. In particular, the span of the first k columns or first k rows of U is uniformly-distributed in $G_{n,k}$.

3. *Fix $x, y \in S^{n-1}$ with $x \perp y$. Then the law of Uy conditioning on $Ux = u$ is uniform in the $(n-2)$-dimensional sphere $u^\perp \cap S^{n-1}$.*

For a subspace $E \subseteq \mathbb{R}^n$, we write $Proj_E$ for the orthogonal projection operator onto $E \subseteq \mathbb{R}^n$. When we project a given point $\theta \in \mathbb{R}^n$ to a random k-dimensional subspace $E \subseteq \mathbb{R}^n$ that is distributed uniformly over the grassmannian, the Euclidean norm of θ decays by a factor of roughly $\sqrt{k/n}$, as is proven in the following:

Lemma 3.7. *Let $1 \leq k \leq n$ and $x_0 \in S^{n-1}$ be fixed, and let $E \in G_{n,k}$ be a random, uniformly distributed subspace. Then for any $\varepsilon > 0$, with probability of at least $1 - C\exp(-c\varepsilon^2 k)$ of selecting E,*

$$(1-\varepsilon)\sqrt{\frac{k}{n}} \leq |Proj_E x_0| \leq (1+\varepsilon)\sqrt{\frac{k}{n}},$$

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

Proof. Rather than having the point x_0 fixed and the subspace E random, we can switch! Let $E_0 \in G_{n,k}$ be a fixed subspace, and let $X \sim \text{Unif}(S^{n-1})$. We claim that

$$|Proj_E x_0| \stackrel{(d)}{=} |Proj_{E_0} X|, \tag{21}$$

i.e., these two random variables are equidistributed. Indeed, if $U \in O(n)$ is a random orthogonal transformation, then X has the same distribution as $U(x_0)$, hence $|Proj_{E_0} X|$ is equidistributed with

$$|Proj_{E_0} U x_0| = |U^* Proj_{E_0} U x_0| = |Proj_{U^* E_0} x_0| \stackrel{(d)}{=} |Proj_E x_0|,$$

proving (21). Our goal is thus to show that $|Proj_{E_0} X|$ is concentrated. For $x \in S^{n-1}$, denote

$$f(x) = |Proj_{E_0} x|.$$

The function f is clearly 1-Lipschitz. Let $u_1, \dots, u_k \in E_0$ be an orthonormal basis, and let us complete it to a full orthonormal basis $u_1, \dots, u_n \in \mathbb{R}^n$. Then

$$\begin{aligned}\int_{S^{n-1}} f^2 d\sigma_{n-1} &= \int_{S^{n-1}} \sum_{i=1}^k (x \cdot u_i)^2 d\sigma_{n-1}(x) = k \cdot \int_{S^{n-1}} (x \cdot u_1)^2 d\sigma_{n-1}(x) \\ &= \frac{k}{n} \int_{S^{n-1}} \sum_{i=1}^n (x \cdot u_i)^2 d\sigma_{n-1}(x) = \frac{k}{n}.\end{aligned}$$

Hence $\sqrt{k/n}$ is a central value of the non-negative, 1-Lipschitz function f, and we may apply Theorem 3.1 with $L = 1$ and with the average E replaced by $\sqrt{k/n}$. We obtain that

$$\sigma_{n-1}\left(\left\{x \in S^{n-1}\,;\, \left|f(x) - \sqrt{\frac{k}{n}}\right| \geq \varepsilon\sqrt{\frac{k}{n}}\right\}\right) \leq Ce^{-\varepsilon^2 k},$$

completing the proof. □

Proof of Lemma 3.5. Denote $m = \#(X)$. Consider the set

$$Y = \left\{\frac{x-y}{|x-y|}\,;\, x \neq y, x, y \in X\right\},$$

so that $\#(Y) \leq m^2$. Let $E \in G_{n,k}$ be a random subspace, uniformly distributed in the Grassmannian. We claim that

$$\mathbb{P}\left(\forall \theta \in Y,\ (1-\varepsilon)\sqrt{\frac{k}{n}} \leq |Proj_E \theta| \leq (1+\varepsilon)\sqrt{\frac{k}{n}}\right) \geq \frac{1}{2}. \tag{22}$$

Indeed, by our choice of parameters, we may ensure that

$$C \exp(-c\varepsilon^2 k) \leq \frac{1}{2m^2}$$

with $c, C > 0$ being the universal constants from Lemma 3.7. Hence, by the *union bound* and the conclusion of that lemma, with probability of at least $1/2$, for all $\theta \in Y$,

$$(1-\varepsilon)\sqrt{\frac{k}{n}} \leq |Proj_E \theta| \leq (1+\varepsilon)\sqrt{\frac{k}{n}}. \tag{23}$$

This proves (22). Since the probability estimated in (22) is positive, we may fix a subspace $E \in G_{n,k}$ for which the event described in (22) holds true. Setting

$$f = \sqrt{\frac{n}{k}} Proj_E,$$

we have that $f : \mathbb{R}^n \to E \cong \mathbb{R}^k$, and for any $x, y \in X$ with $x \neq y$,

$$|f(x) - f(y)| = |x-y| \cdot \left| f\left(\frac{x-y}{|x-y|}\right)\right| = |x-y| \cdot \sqrt{\frac{n}{k}} \left| Proj_E \left(\frac{x-y}{|x-y|}\right)\right|.$$

Since the point $\theta = (x-y)/|x-y|$ belongs to the set Y, from (23) we conclude that

$$1 - \varepsilon \leq \frac{|f(x) - f(y)|}{|x-y|} \leq 1 + \varepsilon,$$

completing the proof. □

We remark that by using entropy considerations, it was shown by Larsen and Nelson [7] that the estimates in the Johnson-Lindenstrauss lemma are close to optimal if ε is not extremely small. They asked whether with $m = \#(X)$, the requirement $k \geq c\varepsilon^{-2} \cdot \log m$ in the Johnson-Lindenstrauss lemma may be improved to $k \geq c\varepsilon^{-2} \cdot \log(2 + \varepsilon^2 m)$. This is still open.

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 3.4.

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

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

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 3.1 by $\sqrt{\int_{S^{n-1}} f^2 d\sigma_{n-1}}$.
(some jargon: any a with $|a - E| \lesssim L/\sqrt{n}$ may be called a "central value" of f).

5. Recall the proof of Haar's theorem, and prove Claim 3.6.

References

[1] Bakry, D., Gentil, I., Ledoux, M., *Analysis and Geometry of Markov Diffusion Operators.* Springer, Berlin, 2014.

[2] 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.

[3] Bobkov, S. G., Chistyakov, G., Götze, F., *Concentration and Gaussian approximation for randomized sums*. Probability Theory and Stochastic Modelling, Vol. 104, Springer, 2023.

[4] Diaconis, P., Freedman, D., *Asymptotics of graphical projection pursuit.* Ann. Statist., 12, no. 3, (1984), 793–815.

[5] Johnson, W. B., Lindenstrauss, J., *Extensions of Lipschitz mappings into a Hilbert space.* Conference in modern analysis and probability, Contemp. Math., Vol. 26, Amer. Math. Soc., (1984), 189–206.

[6] Klartag, B., Lehec, J., *Thin-shell bounds via parallel coupling*. Preprint, arXiv: 2507.15495

[7] Larsen, K. G., Nelson, J., *Optimality of the Johnson-Lindenstrauss lemma.* Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, (2017), 633–638.

[8] Müller, C., *Spherical Harmonics.* Lecture Notes in Math., Vol. 17, Springer, 1966.

[9] Sudakov, V. N., *Typical distributions of linear functionals in finite-dimensional spaces of high-dimension.* (Russian) Dokl. Akad. Nauk. SSSR, 243, no. 6, (1978), 1402–1405. English translation in Soviet Math. Dokl., 19, (1978), 1578–1582.

Thanks to Georgios Gavrilopoulos for correcting typos in an earlier version of these notes.

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.

October 15, 2025