MATH 7251: High-Dimensional Probability Lecture Notes

Galyna V. Livshyts

Scribes: Max Dabagia, Ruijia Cao, Dekang Meng, Xuanzhou Chen, Miao Li, Wanying Fu, Rakshit Naidu, Hanyang Jiang, Rahul Sethi, Chengyue Huang, Yuzhou Wang, William Held, Sam van der Poel, Akash Harapanahalli, Jingyi Zhang, Chen Lin, Daqian Bao, Zichong Li, Adam Brown, Oscar Hasburn-Babich, Shikun Liu, José R. Tuirán Rangel, Yinan Huang, Tian-Yi Zhou, Venkatakrishnan Vaidyanathapuram Krishnan, Jonghyeok Lee, Mengyu Yang

Abstract

Proofreading draft. One author query remains in the convex-body transport homework: whether the intended conclusion concerns contraction or global Lipschitz continuity.

Disclaimer: these lecture notes are currently under construction, and may not be fully proofread yet. If you spot a typo, please let me know! Also, references are in the process of being matched to the text.

Contents

1 Notation and Preliminaries **4**
- 1.1 Notation 4
- 1.2 Preliminaries from high-dimensional geometry and convexity 4
- 1.3 Preliminaries from linear algebra 5
- 1.4 Preliminaries from Functional Analysis and Probability 6

2 The Probabilistic Method and the concept of high-dimensional phenomena **7**
- 2.1 A cool fact about the cube 7
- 2.2 Random rounding 8
- 2.3 Proof of the cool fact about the cube 9

2.4 The Randomized Carathéodory theorem . . . 10
2.5 The standard ε-net argument . . . 13
2.6 A lattice net . . . 16
2.7 An application: an efficient net for matrix multiplication . . . 20
2.8 High-dimensional phenomenon: some themes . . . 23

3 Concentration Inequalities for sums of independent random variables 25
3.1 What is a concentration inequality? . . . 25
3.2 Hoeffding's inequality . . . 26
3.3 Chernoff's inequality . . . 29
3.4 Application to Random Graphs . . . 30
3.5 Sub-Gaussian Random Variables . . . 31
3.6 General Hoeffding's inequality and Khintchine's inequality . . . 35
3.7 Sub-Exponential Random Variables . . . 37
3.8 Bernstein's inequality . . . 39
3.9 Concentration of the norm of a random vector with independent sub-Gaussian coordinates . . . 41
3.10 Sub-Gaussian random vectors . . . 42
3.11 Grothendieck's inequality . . . 44

4 Random Matrices 46
4.1 Norm of a sub-Gaussian random matrix . . . 47
4.2 Two-sided singular-value bounds for tall matrices . . . 49
4.3 Matrix Bernstein Inequality . . . 51
4.4 Non-asymptotic bounds for the smallest singular value of random matrices . 54
4.4.1 General discussion about the smallest singular value of a random matrix 54
4.4.2 Small-ball assumptions and tensorization . . . 55
4.4.3 The smallest singular value of tall random matrices . . . 58
4.4.4 A net construction which works with high probability for matrices with independent columns . . . 60
4.5 Proof of Theorem 4.25. . . . 60
4.6 The smallest singular value of square random matrices . . . 66
4.6.1 Rudelson–Vershynin decomposition of the sphere . . . 68
4.6.2 Survey of results regarding the smallest singular value of square random matrices . . . 69
4.7 Proof of Theorem 4.45 Part 1 . . . 72
4.7.1 Compressible Vectors . . . 72
4.7.2 Incompressible Vectors . . . 73
4.7.3 Distance Theorem . . . 75
4.7.4 Proof of the first part of Theorem 4.45 . . . 77

5 Gaussian Random Processes **78**
5.1 Basic Concepts and Examples 78
5.2 Slepian's Inequality 80
5.2.1 Gaussian Interpolation 81
5.2.2 Proof of Slepian's Inequality 83
5.2.3 The Sudakov-Fernique Inequality 85
5.2.4 Application of Sudakov-Fernique Inequality to Gaussian Random Matrices 85

6 The Semigroup method **88**
6.1 Basic definitions and set up 88
6.2 Properties of Markov semigroups, and some examples 90
6.3 The Ornstein–Uhlenbeck semigroup 96
6.4 Gaussian Poincare inequality via the semigroup method 98
6.5 A discussion on Poincare inequalities, the example of the circle and periodic functions 99
6.6 Proof of the abstract Poincare-type inequality for stationary measures of semigroups 101
6.7 The Gaussian Log-Sobolev Inequality of Gross via the semigroup method . . 104
6.8 Gaussian isoperimetry via the semigroup method 106
6.9 Bobkov's Inequality via the semigroup method 108
6.9.1 Semigroup proof of Bobkov's inequality 109
6.9.2 Applications to concentration of measure 111

7 The mass transport method **118**
7.1 Basic definitions and setup 118
7.2 Useful Properties of Lipschitz Transport Maps 118
7.3 Optimal Transport Map with respect to quadratic cost 121
7.4 Monge Problem 124
7.5 Kantorovich Problem 125
7.6 Cyclic Monotonicity 125
7.7 Kantorovich Duality 127
7.7.1 Examples 127
7.8 Brenier's Theorem 129
7.9 Mass transport proof of the Brunn-Minkowski inequality 130
7.10 Log-concave measures 134
7.11 Applications of Brenier's theorem in HDP; Caffarelli's theorem 134
7.12 The Talagrand transport inequality 137

1 Notation and Preliminaries

1.1 Notation

- $\mathbb{R}^n$ is the n-dimensional Euclidean space
- Lebesgue measure (volume) in $\mathbb{R}^n$ of a measurable set $A \subset \mathbb{R}^n$ is denoted by $|A|$
- $\mathbb{N}$ is the set of positive integers
- $\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$ is the p-norm in $\mathbb{R}^n$ for $1 \leq p < \infty$
- $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$ is the ∞-norm
- $|x| = \|x\|_2$ is a shorthand for Euclidean length
- $\mathbf{B}_p^n = \{x \in \mathbb{R}^n \colon \|x\|_p \leq 1\}$ is the unit p-ball in $\mathbb{R}^n$
- $\mathbb{S}^{n-1} = \partial \mathbf{B}_2^n = \{x \in \mathbb{R}^n \colon |x| = 1\}$ is the $(n-1)$-dimensional unit sphere in $\mathbb{R}^n$, or the boundary of $\mathbf{B}_2^n$
- $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ is the standard inner product
- For $\theta \in \mathbb{S}^{n-1}$, we have the hyperplane

  $$\theta^\perp = \{x \in \mathbb{R}^n \colon \langle x, \theta \rangle = 0\}$$

  and the affine hyperplane

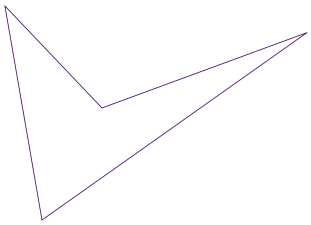
  $$\theta^\perp + t\theta = \{x \in \mathbb{R}^n \colon \langle x, \theta \rangle = t\}$$

  for all $t \in \mathbb{R}$.
- A half-space is a set of the form $\{x \in \mathbb{R}^n : \langle x, \theta \rangle \leq t\}$, for $\theta \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}$.
- A strip is a set of the form $\{x \in \mathbb{R}^n : |\langle x - y, \theta \rangle| \leq t\}$, for $y \in \mathbb{R}^n$, $\theta \in \mathbb{S}^{n-1}$ and $t \geq 0$.
- Fix $x \in \mathbb{R}$. Then $[x]$ is the floor function, the largest integer which is no larger than x.

1.2 Preliminaries from high-dimensional geometry and convexity

Definition 1.1 (Convex Set)**.** A set $K \subseteq \mathbb{R}^n$ is called *convex* if for all $x, y \in K$, the line segment $[x, y] = \{\lambda x + (1-\lambda)y : \lambda \in [0,1]\}$ is contained in K.

Note that strips, half-spaces and B_p^n for $p \geq 1$ are convex sets. On the other hand, B_p^n for $0 < p < 1$ and $n \geq 2$, $\mathbb{S}^{n-1}$, and nonempty sets that are not simply connected are non-convex. Below is an example of a non-convex set:



Any closed convex set is the intersection of (possibly infinitely many) closed half-spaces. An intersection of finitely many closed half-spaces is called a convex *polyhedron*; a bounded polyhedron is a convex *polytope*.

1.3 Preliminaries from linear algebra

For a real $N \times n$ matrix A, the operator norm $\|\cdot\|_{op}$ is defined by

$$\|\mathbf{A}\|_{op} = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|\mathbf{A}x|}{|x|} = \sup_{y \in \mathbb{S}^{n-1}} |Ay|. \tag{1}$$

Definition 1.2 (Hilbert-Schmidt norm)**.** Given a real $N \times n$ matrix $\mathbf{A} = (a_{ij})$,

$$\|\mathbf{A}\|_{HS} = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\sigma_1^2 + ... + \sigma_n^2},$$

where $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ are the square roots of the eigenvalues of $\mathbf{A}^T\mathbf{A}$, including zeros. This convention also covers rectangular matrices with fewer rows than columns.

Recall that the smallest singular value is also defined as

$$\sigma_n(A) = \inf_{x \in \mathbb{S}^{n-1}} |Ax|.$$

Remark 1.3. *Note that*

$$\sqrt{n}\|\mathbf{A}\|_{op} \geq \|\mathbf{A}\|_{HS} \geq \|\mathbf{A}\|_{op},$$

since $\sigma_i \leq \sigma_1$*, and* $\sigma_1 \leq \sqrt{\sigma_1^2 + ... + \sigma_n^2} \leq \sqrt{n}\sigma_1$*.*

Theorem 1.4 (Spectral Decomposition)**.** *Let A be a real symmetric $n \times n$ matrix. It has real eigenvalues $\lambda_1, \ldots, \lambda_n$ and a corresponding orthonormal basis of eigenvectors $u_1, \ldots, u_n \in \mathbb{S}^{n-1}$. Then*

$$A = \sum_{i=1}^{n} \lambda_i u_i \otimes u_i \tag{2}$$

For every $u \in \mathbb{R}^n$, the matrix $u \otimes u = uu^T = \begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_1u_2 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \vdots & \vdots \\ u_1u_n & u_2u_n & \cdots & u_nu_n \end{bmatrix}$ has rank one when $u \neq 0$, and rank zero when $u = 0$. For every $x \in \mathbb{R}^n$, $Ax = \sum_{i=1}^n \lambda_i \langle x, u_i \rangle u_i$.

Definition 1.5 (Functions on Matrices)**.** For any function $f : \mathbb{R} \to \mathbb{R}$ and $n \times n$ symmetric matrix $X = \sum_{i=1}^{n} \lambda_i u_i \otimes u_i$, where the u_i form an orthonormal eigenbasis, we define the matrix $f(X) := \sum_{i=1}^{n} f(\lambda_i) u_i \otimes u_i$. This definition does not depend on the choice of orthonormal eigenbasis.

1.4 Preliminaries from Functional Analysis and Probability

Recall that for any non-negative random variable

$$\mathbb{E}\, X = \int_0^\infty \mathbb{P}(X > t) dt. \tag{3}$$

Indeed, $X = \int_0^\infty 1_{\{X>t\}} dt$, and therefore

$$\mathbb{E}\, X = \mathbb{E} \int_0^\infty 1_{\{X>t\}} dt = \int_0^\infty \mathbb{E}\, 1_{\{X>t\}} dt = \int_0^\infty \mathbb{P}(X > t) dt.$$

Lemma 1.6 (Markov)**.** *Let $X \geq 0$ be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for $t > 0$, we have*

$$\mathbb{P}\left(X \geq t\right) \leq \frac{\mathbb{E}\, X}{t}.$$

Proof. For any $t > 0$, note that

$$\mathbb{E}\, X = \int_0^\infty \mathbb{P}(X > s)\ \mathrm{d}s \geq \int_0^t \mathbb{P}(X > s)\ \mathrm{d}s \geq t \cdot \mathbb{P}(X \geq t).$$

□

Definition 1.7. A function $F\colon \mathbb{R}^n \to \mathbb{R}$ is called **convex** if for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$

$$F(\lambda x + (1 - \lambda) y) \leq \lambda F(x) + (1 - \lambda) F(y).$$

Remark 1.8. *Note that if F is convex, then we may conclude inductively that*

$$F\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{1 \leq i \leq m} \lambda_i F(x_i),$$

where $\lambda_i \geq 0$ and $\lambda_1 + \cdots + \lambda_m = 1$.

Theorem 1.9 (Jensen Inequality)**.** *Let μ be a probability measure on $\mathbb{R}^n$ and let $g\colon \mathbb{R}^n \to \mathbb{R}^m$ have integrable coordinates with respect to μ. Let $F\colon \mathbb{R}^m \to \mathbb{R}$ be a convex function. Then,*

$$F\left(\int_{\mathbb{R}^n} g\ \mathrm{d}\mu\right) \leq \int_{\mathbb{R}^n} F(g)\ \mathrm{d}\mu.$$

Remark 1.10. *Note that Jensen's inequality implies Remark 1.8 by taking* $\mu = \sum_{i=1}^{m} \lambda_i \delta_{x_i}$ *and* $g(x) = x$.

Definition 1.11. Let X be a random variable. For $1 \leq p < \infty$, the p-norm of X is defined by

$$\|X\|_p = \left(\mathbb{E}\,|X|^p\right)^{1/p}.$$

Theorem 1.12 (Minkowski)**.** *Let* X, Y *be two random variables on the same probability space and let* $1 \leq p < \infty$. *Then,*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Theorem 1.13 (Cauchy-Schwarz)**.** *Let* $a, b \in V$, *where* $(V, \langle \cdot, \cdot \rangle)$ *is an inner product space. Then,*

$$|\langle a, b \rangle| \leq \|a\| \cdot \|b\|,$$

where $\|\cdot\|$ *denotes the norm induced by* $\langle \cdot, \cdot \rangle$.

Theorem 1.14 (Hölder's Inequality)**.** *Suppose that* $1 < p, q < \infty$ *and* $1/p + 1/q = 1$. *Then,*

$$\int |fg| \, \mathrm{d}\mu \leq \left(\int |f|^p \, \mathrm{d}\mu\right)^{1/p} \cdot \left(\int |g|^q \, \mathrm{d}\mu\right)^{1/q}$$

The endpoint case is $\int |fg| \, \mathrm{d}\mu \leq \|f\|_{L^1(\mu)} \|g\|_{L^\infty(\mu)}$, *and similarly with* f *and* g *interchanged.*

2 The Probabilistic Method and the concept of high-dimensional phenomena

2.1 A cool fact about the cube

We will mostly study $\mathbb{R}^n$ – the n-dimensional Euclidean space – where n is a large, positive integer (or, in other words, the dimension n of our space is high). We will often think that $n \to \infty$ and analyze things asymptotically. We know intuitively that functions on $\mathbb{R}^2$ are more complex than those on $\mathbb{R}$, and functions on $\mathbb{R}^3$ are more complex than on $\mathbb{R}^2$. However, in some sense, objects in high-dimensional spaces actually become simpler, and more predictable. We will soon see some examples of this phenomenon, but for the time being we concentrate on the so-called probabilistic method in High-Dimensional Geometry.

We now state the following cool fact:

Fact 2.1. *Consider the cube* $[0,1]^n$ *(with side-length* 1*) and pick* $x \in [0,1]^n$. *Take any (arbitrary!)* $\theta \in \mathbb{S}^{n-1}$, *and consider the strip centered at* x *orthogonal to* θ *of width* 1, *i.e.*

$$S = \left\{ y \in \mathbb{R}^n \colon |\langle y - x, \theta \rangle| \leq \frac{1}{2} \right\}.$$

Then at least one vertex of $[0,1]^n$ *belongs to* S.

The proof will be based on the *probabilistic method*. Imagine that someone shows you a non-transparent box full of balls and tells you that if you draw a ball from it then with probability 0.3 you get a red ball from it. Then you can conclude that *there exists* at least one red ball in the box.



In our case, the box will be the vertices of the cube, and the red ball will be the vertex with the desired property of falling into the specific strip.

2.2 Random rounding

We will first define the concept of randomized rounding, which was developed as an algorithmic technique by Raghavan and Thompson [32], and later extended and studied by many authors including Kannan and Vempala [17], Alon and Klartag [2], Klartag and Livshyts [19], see a survey by Srinivasan [44]. This object is very useful in Computer Science, as well as in High-Dimensional Probability and related areas.

Definition 2.2 (Random Rounding)**.** We outline the definition in two steps.

Step 1 (dimension 1). For $x \in \mathbb{R}$, define η_x to be a random variable such that

$$\eta_x = \begin{cases} [x] & \text{w.p. } 1-p \\ [x]+1 & \text{w.p. } p, \end{cases}$$

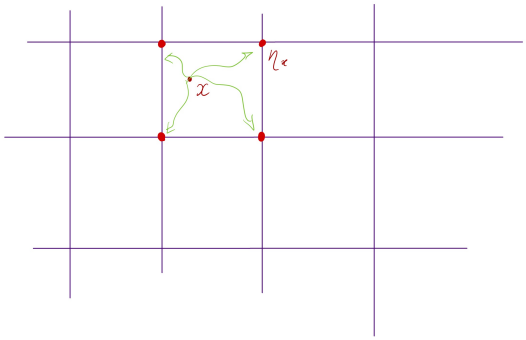
where $p = p(x) = x - [x]$. Note that p is chosen so that η_x is centered at x:

$$\mathbb{E}\,\eta_x = [x](1-(x-[x])) + ([x]+1)(x-[x]) = [x] + x - [x] = x.$$

Step 2 (dimension n). For $x \in \mathbb{R}^n$, define η_x to be a random vector taking values in the vertex set of the lattice cube which x falls into, so that the coordinates of η_x are independent and $\mathbb{E}\,\eta_x = x$. In other words, each coordinate of x is independently randomly rounded using the 1-dimensional definition. Namely, we let $\eta_x = ((\eta_x)_1, ..., (\eta_x)_n)$, where the random variables $(\eta_x)_i$ are independent, and distributed as follows:

$$(\eta_x)_i = \begin{cases} [x_i] & \text{w.p. } 1-p_i \\ [x_i]+1 & \text{w.p. } p_i, \end{cases}$$

where $p_i = x_i - [x_i]$.



2.3 Proof of the cool fact about the cube

First, recall the following basic fact:

Lemma 2.3. *For independent random variables $X_1, \dots, X_n$ with finite second moments,*

$$\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n).$$

Proof. First, consider

$$\mathbb{E}(X_1 + ... + X_n)^2 = \sum_{i=1}^n \mathbb{E}\, X_i^2 + 2\sum_{i<j} \mathbb{E}\, X_i X_j = \sum_{i=1}^n \mathbb{E}\, X_i^2 + 2\sum_{i<j} \mathbb{E}\, X_i\, \mathbb{E}\, X_j,$$

where in the last step we used independence, $\mathbb{E}\, X_i X_j = \mathbb{E}\, X_i\, \mathbb{E}\, X_j$. On the other hand,

$$(\mathbb{E}\, X_1 + ... + \mathbb{E}\, X_n)^2 = \sum_{i=1}^n (\mathbb{E}\, X_i)^2 + 2\sum_{i<j} \mathbb{E}\, X_i\, \mathbb{E}\, X_j$$

Therefore,

$$\begin{aligned}\operatorname{Var}(X_1 + \cdots + X_n) &= \mathbb{E}(X_1 + \cdots + X_n)^2 - (\mathbb{E}(X_1 + \cdots + X_n))^2 \\ &= \sum_{i=1}^n \left(\mathbb{E}\, X_i^2 - (\mathbb{E}\, X_i)^2\right) = \sum_{i=1}^n \operatorname{Var}(X_i).\end{aligned}$$

□

We may now proceed with the proof of the cool fact.

Proof of Fact 2.1. For the vector $x \in [0,1]^n$ consider the random rounding η_x. Note that $\mathbb{E}\langle \eta_x - x, \theta\rangle = 0$, so

$$\mathbb{E}\langle \eta_x - x, \theta\rangle^2 = \operatorname{Var}\langle \eta_x - x, \theta\rangle = \sum_{i=1}^n \mathbb{E}((\eta_x)_i - x_i)^2 \theta_i^2,$$

by Lemma 2.3. Hence,

$$\mathbb{E}\langle \eta_x - x, \theta\rangle^2 = \sum_{i=1}^{n} \theta_i^2 (\mathbb{E}(\eta_x)_i^2 - x_i^2).$$

Note that (by definition of random rounding),

$$\mathbb{E}(\eta_x)_i^2 = 0 \cdot (1 - x_i) + 1 \cdot x_i = x_i,$$

so

$$\mathbb{E}(\eta_x)_i^2 - x_i^2 = x_i - x_i^2 \leq \frac{1}{4},$$

where we use the fact that for any $a \in [0,1]$ one has $a - a^2 \leq \frac{1}{4}$. We conclude that

$$\mathbb{E}\langle \eta_x - x, \theta\rangle^2 = \sum_{i=1}^{n} \theta_i^2 x_i(1 - x_i) \leq \frac{|\theta|^2}{4} = \frac{1}{4}. \tag{4}$$

Therefore, there exists some realization of η_x, a vertex $y \in \{0,1\}^n$ such that $\langle y - x, \theta\rangle^2 \leq 1/4$. Hence, $|\langle y - x, \theta\rangle| \leq 1/2$. □

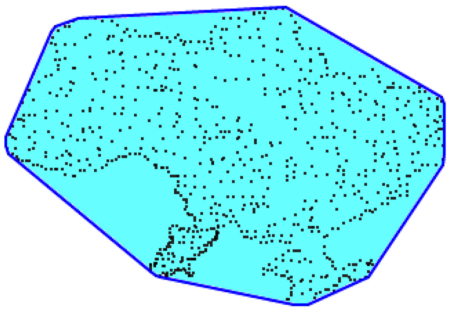
2.4 The Randomized Carathéodory theorem

We will now state a second cool fact, which is an approximate version of Carathéodory's theorem. We follow Vershynin [55] in this subsection.

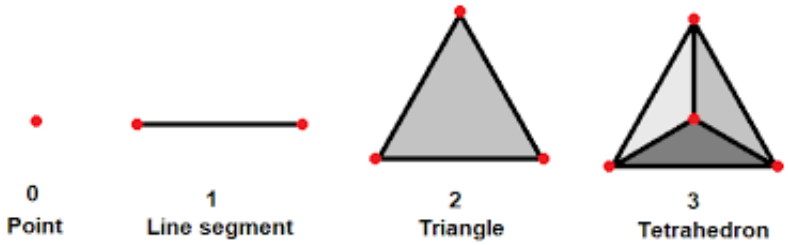
Definition 2.4 (Convex hull)**.** Let $A \subseteq \mathbb{R}^n$. The convex hull of A is

$$\mathrm{conv}(A) = \left\{ \sum_{i=1}^{m} \lambda_i x_i : \; m \in \mathbb{N},\, x_1, \ldots, x_m \in A,\, \lambda_i \geq 0,\, \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

The expression $\sum_{i=1}^{m} \lambda_i x_i$ is called a *convex combination* of $x_1, ..., x_m$.



For example, the convex hull of $n+1$ affinely independent points in $\mathbb{R}^n$ is called a *simplex*.



The next classical result tells us that every point in a convex hull of some set in $\mathbb{R}^n$ can be represented as a convex combination of at most $n+1$ points from this set, i.e. m in the definition of convex hull need not be larger than $n+1$.

Theorem 2.5 (Carathéodory's theorem)**.** *Let $A \subseteq \mathbb{R}^n$ and $x \in \text{conv}(A)$. Then there exist $x_1, \ldots, x_{n+1} \in A$ and $\lambda_1, \ldots, \lambda_{n+1} \geq 0$, $\sum_{i=1}^{n+1} \lambda_i = 1$ such that $x = \sum_{i=1}^{n+1} \lambda_i x_i$. In other words, x belongs to a simplex of dimension at most n whose vertices lie in A.*

Proof. Homework! □

Carathéodory's Theorem could be useful for constructing various algorithms. However, having to operate with $n+1$ points could still be too difficult if $n = 100000000$, say. Would it be possible to represent x with fewer than $n+1$ points from A? In general, of course, not. However, if we were willing to represent x *approximately*, with some small error, then we could get away with using far fewer points, potentially:

Fact 2.6 (Randomized/Approximate Carathéodory)**.** *Suppose $A \subseteq \mathbb{R}^n$ with diameter*

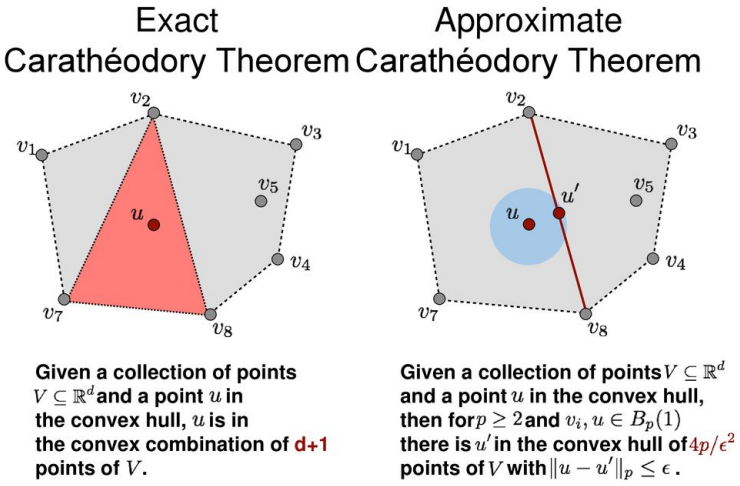
$$\text{diam}(A) = \sup_{x,y \in A} |x - y| \leq 1.$$

Then for all $x \in \text{conv}(A), k \in \mathbb{N}$, there exist $x_1, \ldots, x_k \in A$, such that

$$\left\| x - \frac{1}{k} \sum_{i=1}^{k} x_i \right\| \leq \frac{1}{\sqrt{k}}.$$

Note that we allow repetitions among the points.

Although $\frac{1}{k} \sum_{i=1}^{k} x_i$ need not equal x, it approximates x with error at most $1/\sqrt{k}$. The advantage is that we only need k points, and potentially $k \ll n+1$. Another interesting feature here is that we can take $\lambda_1 = \lambda_2 = \ldots = \lambda_k = \frac{1}{k}$, rather than deal with different weights.



Proof of Fact 2.6. Choose $x \in \mathrm{conv}(A)$. By Carathéodory's theorem, there are points $z_1, \ldots, z_{n+1} \in A$ and nonnegative coefficients λ_i with $\sum_{i=1}^{n+1} \lambda_i = 1$ such that

$$x = \sum_{1 \leq i \leq n+1} \lambda_i z_i.$$

Consider a random vector Z which takes value z_1 with probability λ_1, value z_2 with probability λ_2, and so on (overall, this random vector takes values among $z_1, \ldots, z_{n+1} \in A$). Note that

$$\mathbb{E}\, Z = \sum_{1 \leq i \leq n+1} \lambda_i z_i = x.$$

Now, given an integer $k \geq 1$, consider k i.i.d. copies of Z denoted by $Z_1, \ldots, Z_k$. They can be constructed on the k-fold product of the probability space supporting Z. Note that by Lemma 2.3,

$$\mathbb{E} \left| x - \frac{1}{k} \sum_{1 \leq i \leq k} Z_i \right|^2 = \frac{1}{k^2} \sum_{1 \leq i \leq k} \mathbb{E}\, |x - Z_i|^2 \leq \frac{1}{k^2} \cdot k = \frac{1}{k},$$

The equality follows by expanding the squared Euclidean norm: all mixed terms vanish because the $Z_i - x$ are independent and have mean zero. For the inequality, each Z_i belongs to A and $x \in \mathrm{conv}(A)$, so $|x - Z_i| \leq 1$. Indeed, $\mathrm{diam}(\mathrm{conv}(A)) = \mathrm{diam}(A) \leq 1$ (homework!).

We conclude that

$$\mathbb{E} \left| x - \frac{1}{k} \sum_{1 \leq i \leq k} Z_i \right|^2 \leq \frac{1}{k}.$$

Hence, there exist $y_1, \ldots, y_k \in A$, realizations of $Z_1, \ldots, Z_k$, such that

$$\left| x - \frac{1}{k} \sum_{1 \le j \le k} y_j \right| \le \frac{1}{\sqrt{k}},$$

which proves our Fact. In fact, if Z' is an independent copy of Z, then

$$\mathbb{E}\,|Z - x|^2 = \frac{1}{2}\,\mathbb{E}\,|Z - Z'|^2 \le \frac{1}{2}.$$

Thus the same argument gives the slightly sharper error bound $1/\sqrt{2k}$, which is useful in the next corollary. □

The Approximate Carathéodory Theorem has, for example, the following useful consequence:

Corollary 2.7. *Let P be a nonempty polytope in $\mathbb{R}^n$ with at most $N \ge 1$ vertices, i.e.,*

$$P = \operatorname{conv}\{x_1, \ldots, x_N\},$$

where each $x_i \in \mathbb{R}^n$. Suppose that $\operatorname{diam}(P) \le 1$. Fix any $\varepsilon > 0$. Then, P can be covered by at most $N^{\lfloor 1/\varepsilon^2 \rfloor}$ Euclidean balls of radius ε.

Proof. Homework! For $0 < \varepsilon < 1$, use the sharper bound $1/\sqrt{2k}$ with $k = \lfloor 1/\varepsilon^2 \rfloor \ge 1/(2\varepsilon^2)$, and count the possible averages of k vertices. For $\varepsilon \ge 1$, one ball suffices. □

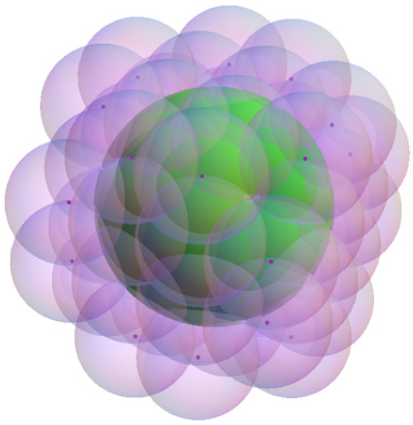
While we leave this fact as homework, in the next subsection we prove a related and very useful result.

2.5 The standard ε-net argument

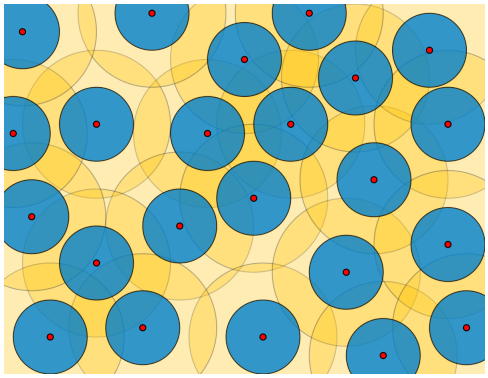
Suppose we would like to view the sphere or the unit ball as a discrete set, in which each point is represented approximately. What would be a good way to do that? We start by presenting the most classical result.

Theorem 2.8 (Classical ε-net Construction)**.** *Let $n \in \mathbb{N}$ and $\varepsilon > 0$. Then there exist a positive integer $m \le ((2+\varepsilon)/\varepsilon)^n$ and points $y_1, \ldots, y_m \in \mathbf{B}_2^n$ such that the unit ball $\mathbf{B}_2^n$ is covered by balls of radius ε with the centers at y_i, that is,*

$$\mathbf{B}_2^n \subseteq \bigcup_{1 \le i \le m} (y_i + \varepsilon \mathbf{B}_2^n).$$



Proof. Without loss of generality, we may assume that $\varepsilon < 1$ (otherwise, the statement is straightforward). Consider a **maximal** packing of balls of radius $\varepsilon/2$ with centers in $\mathbf{B}_2^n$: pick $y_1, \dots, y_m \in \mathbf{B}_2^n$ such that the closed balls $y_i + (\varepsilon/2)\mathbf{B}_2^n$ are disjoint and no further such ball with center in $\mathbf{B}_2^n$ can be added. All these balls lie in $(1+\varepsilon/2)\mathbf{B}_2^n$. Here maximal means maximal under adding another center, not necessarily of largest cardinality (see the remarks below for existence).



First, we show that

$$B_2^n \subseteq \bigcup_{1 \le i \le m} (y_i + \varepsilon B_2^n). \tag{5}$$

Suppose not. Then, there exists $x \in B_2^n$ such that $|x - y_i| > \varepsilon$ for all $1 \le i \le m$, and $x + \frac{\varepsilon}{2}\mathbf{B}_2^n \subset (1+\frac{\varepsilon}{2})\mathbf{B}_2^n$. But then $(x + (\varepsilon/2)\mathbf{B}_2^n) \cap (y_i + (\varepsilon/2)\mathbf{B}_2^n) = \emptyset$ for every i, contradicting maximality. Hence (5) is confirmed. That is, we found a construction of some covering of B_2^n with balls of radius ε; our next task is to estimate its size.

Observe that

$$(1 + \varepsilon/2)\mathbf{B}_2^n \supseteq \bigsqcup_{1 \le i \le m} (y_i + \varepsilon/2\mathbf{B}_2^n)$$

Therefore, by the additive property of the Lebesgue measure, we have

$$\sum_{1 \le i \le m} \left| y_i + \frac{\varepsilon}{2}\mathbf{B}_2^n \right| \le \left| (1 + \varepsilon/2)\mathbf{B}_2^n \right|,$$

where $|\cdot|$ denotes the n-dimensional Lebesgue measure. Therefore, we have

$$m \cdot \left(\frac{\varepsilon}{2}\right)^n \leq \left(1 + \frac{\varepsilon}{2}\right)^n$$

Hence, $m \leq (1 + 2/\varepsilon)^n$, as desired. □

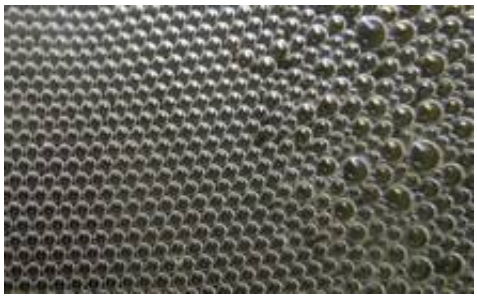
Remark 2.9 (Ideas by Ruijia Cao, not something Galyna suggested)**.** *To show that a maximal packing used in the proof above exists, we can use the following argument. Note that if $t > 0$ and a packing of m disjoint balls satisfies*

$$\bigsqcup_{i=1}^{m} (y_i + t\mathbf{B}_2^n) \subseteq \mathbf{B}_2^n,$$

then the volume comparison gives $m \leq t^{-n}$. Let $\varepsilon > 0$. We will use the following algorithm to explicitly construct a maximal packing of balls of radius $\varepsilon/2$ with centers in $\mathbf{B}_2^n$:

1. *First, pick any $x_1 \in \mathbf{B}_2^n$.*
2. *In the second step, choose $x_2 \in \mathbf{B}_2^n$ such that $\|x_2 - x_1\| > \varepsilon$.*
3. *At the k-th step, suppose that $x_1, \ldots, x_{k-1}$ have been chosen such that $x_i + (\varepsilon/2)\mathbf{B}_2^n$ are pairwise disjoint. Choose $x_k \in \mathbf{B}_2^n$ such that $\|x_k - x_i\| > \varepsilon$ for all $1 \leq i \leq k-1$.*
4. *If no such x_k exists, then the algorithm terminates.*

All selected balls lie in $(1 + \varepsilon/2)\mathbf{B}_2^n$, so the same volume comparison bounds their number by $(1 + 2/\varepsilon)^n$. Thus the algorithm terminates after finitely many steps.



Remark 2.10 (a small addition by Galyna)**.** *Arguing along the lines of the previous remark (by Ruijia), one can also choose a packing of largest cardinality. The possible cardinalities form a nonempty bounded set of positive integers, so their supremum is attained. Such a packing is certainly maximal under adding another ball, which is all the proof needs. For details, see e.g. Vershynin [55].*

Remark 2.11. *Note that the size m of the covering of the unit ball by balls of radius ε is necessarily at least $\frac{1}{\varepsilon^n}$, since $m \cdot |\varepsilon \mathbf{B}_2^n| \geq |\mathbf{B}_2^n|$. Therefore, the bound in Theorem 2.8 is sharp up to 3^n, when $\varepsilon < 1$.*

Therefore, if m is the smallest possible size of an ε-covering and $0 < \varepsilon < 1$, we may write $m = (c/\varepsilon)^n$ with $1 \leq c \leq 3$; the optimal c can depend on both n and ε. There is a vast literature on sharper covering bounds. A related but distinct problem is sphere packing*, where one maximizes the density of disjoint equal balls. The optimal planar packing is assigned as a (difficult!) homework problem. In dimensions 8 and 24, optimal sphere packings were established by Viazovska [57] and by Cohn, Kumar, Miller, Radchenko and Viazovska [58], respectively. These packing results do not determine the sharp covering constants above.*



Remark 2.12. *For $0 < \varepsilon < 1$, since $\mathbb{S}^{n-1} \subset \mathbf{B}_2^n$, Theorem 2.8 implies that the unit sphere has a covering by at most $(3/\varepsilon)^n$ balls of radius ε. In fact, for $n \geq 2$ there is such a cover of size at most $(c/\varepsilon)^{n-1}$, where c is an absolute constant. This important fact is left as homework. For $n = 1$, the sphere consists of two points.*

Theorem 2.8 is very classical and it has many powerful applications, some of which we will see in this course. However, there are situations when this result is, in fact, insufficiently sharp. Another disadvantage of the argument we presented is that it does not give an explicit list of the net points (as one of the students cleverly pointed out): the greedy construction still requires choosing a new center outside all the balls already selected.

2.6 A lattice net

In this subsection, we discuss a refinement of Theorem 2.8 which is furthermore constructive (although of course one cannot only gain and not lose, so some features of Theorem 2.8 might not be kept). First, we note the following classical

Lemma 2.13. *Let $n, N \in \mathbb{Z}$ be positive integers. Then, there are $\binom{N+n-1}{n-1}$ solutions to the following equation*

$$x_1 + \cdots + x_n = N,$$

where $x_i \in \mathbb{Z}$ and $x_i \geq 0$.

Proof. Note that the number of ways to represent N as an ordered sum of n non-negative integers is the same as the number of ways to distribute N identical balls into n different boxes (any box can contain between 0 and N balls). The answer to this classical combinatorial riddle is $\binom{N+n-1}{n-1}$. Indeed, this corresponds to the number of ways to arrange a sequence of N zeroes (corresponding to the balls) and $n-1$ ones (corresponding to the sides of the boxes, except the two on the edges). And this is the same as selecting $n-1$ spots to place ones in the collection of $N+n-1$ spots. See Figure 1. □

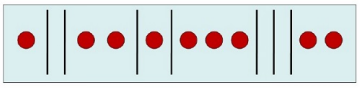


Figure 1: $(n-1)$ dividers and N balls

Theorem 2.14. *Let $R \geq n$. Then there exists an absolute constant $C > 0$ such that*

$$\#\{x \in \mathbb{Z}^n : \|x\|_1 \leq R\} \leq \left(\frac{CR}{n}\right)^n.$$

Proof. Applying Lemma 2.13, and also including the zero vector, we have

$$\begin{aligned}\#\{x \in \mathbb{Z}^n : \|x\|_1 \leq R\} &\leq 2^n \#\{x \in \mathbb{Z}^n : x_i \geq 0,\ \textstyle\sum_{i=1}^n x_i \leq R\} \\ &= 2^n \sum_{N=0}^{\lfloor R \rfloor} \binom{N+n-1}{n-1} = 2^n \binom{\lfloor R \rfloor + n}{n}.\end{aligned}$$

Here the term with $N = 0$ equals 1. Recall Stirling's formula,

$$k! = (1+o(1))\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \qquad \text{as } k \to \infty.$$

The standard estimate $\binom{M}{n} \leq (eM/n)^n$ for $M \geq n$ (homework!) now gives

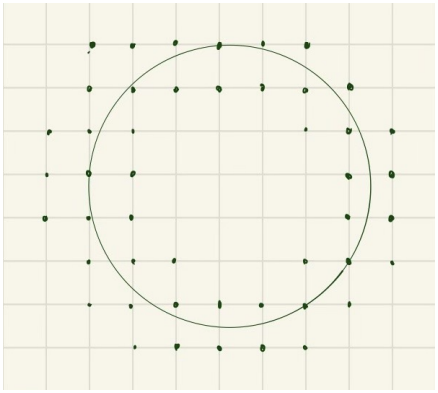
$$2^n \binom{\lfloor R \rfloor + n}{n} \leq \left(\frac{2e(R+n)}{n}\right)^n \leq \left(\frac{4eR}{n}\right)^n,$$

where the last inequality uses $R \geq n$. For arbitrary $R > 0$, the preceding argument instead gives the bound $[2e(1+R/n)]^n$. □

Corollary 2.15. *Fix $0 < \varepsilon \leq 1$. Then there exist $y_1, \ldots, y_m \in \mathbf{B}_2^n$ and an absolute constant $C > 0$ such that*

$$\bigcup_{1 \leq i \leq m} \left(y_i + \frac{\varepsilon}{\sqrt{n}} \mathbf{B}_\infty^n\right) \supseteq \mathbf{B}_2^n,$$

where $m \leq (C/\varepsilon)^n$.



Proof. Let $R = n/\varepsilon \geq n$, and enumerate the lattice points

$$\{z_1, \ldots, z_m\} = \mathbb{Z}^n \cap \frac{R}{\sqrt{n}}\mathbf{B}_2^n.$$

Since $\|z\|_1 \leq \sqrt{n}|z|$, these points satisfy $\|z_i\|_1 \leq R$, so Theorem 2.14 gives $m \leq (C/\varepsilon)^n$.

For any $z \in (R/\sqrt{n})\mathbf{B}_2^n$, round each coordinate *toward zero*:

$$w_j = \operatorname{sgn}(z_j)\lfloor |z_j| \rfloor, \qquad j = 1, \ldots, n,$$

where $\operatorname{sgn}(0) = 0$. Then $|w| \leq |z|$ and $\|z - w\|_\infty < 1$, so w is one of the points z_i. Consequently,

$$\frac{R}{\sqrt{n}}\mathbf{B}_2^n \subseteq \bigcup_{i=1}^m (z_i + \mathbf{B}_\infty^n).$$

Scaling by $\sqrt{n}/R = \varepsilon/\sqrt{n}$ and setting $y_i = (\varepsilon/\sqrt{n})z_i$ gives

$$\mathbf{B}_2^n \subseteq \bigcup_{i=1}^m \left(y_i + \frac{\varepsilon}{\sqrt{n}}\mathbf{B}_\infty^n \right), \qquad y_i \in \mathbf{B}_2^n,$$

as required. Rounding toward zero, rather than taking the floor of a negative coordinate, is what keeps the centers inside the Euclidean ball. □

Remark 2.16. *Here, for $0 < \varepsilon \leq 1$, we compare Theorem 2.8 and Corollary 2.15. By Theorem 2.14 and Corollary 2.15,*

$$\mathbf{B}_2^n \subseteq \bigcup_{1 \leq i \leq m} \left(y_i + \frac{\varepsilon}{\sqrt{n}}\mathbf{B}_\infty^n \right), \tag{6}$$

where $m \leq (\frac{c}{\varepsilon})^n$. On the other hand, Theorem 2.8 implies that

$$\mathbf{B}_2^n \subseteq \bigcup_{1 \leq i \leq m} \left(y_i + \varepsilon \mathbf{B}_2^n \right), \tag{7}$$

where $m \le \left(\frac{3}{\varepsilon}\right)^n$.

Note that in some sense, Corollary 2.15 is stronger than Theorem 2.8 since $\mathbf{B}_\infty^n \subseteq \sqrt{n}\mathbf{B}_2^n$ *(see Figure 2), and therefore (6) implies (7). However, Corollary 2.15 does have a minor disadvantage compared with Theorem 2.8: the constant* $C > 3$ *is unspecified in our computation. But this usually doesn't play a role in applications.*

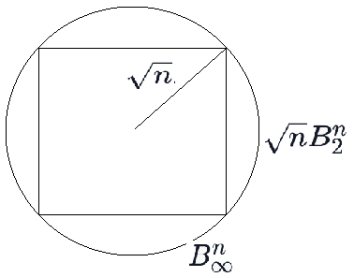


Figure 2: Comparison of the ℓ_2 and ℓ_∞ norms

Since $\mathbb{S}^{n-1} \subset \mathbf{B}_2^n$, we may refine the preceding lattice count by retaining only points in a thin shell near the sphere. This gives the following covering result (the improvement in the exponent requires counting the shell, not just discarding points):

Corollary 2.17 (Homework)**.** *Let* $n \ge 2$ *and* $0 < \varepsilon < 1$. *There exist* $y_1, \dots, y_m \in \mathbf{B}_2^n$ *such that for every* $x \in \mathbb{S}^{n-1}$ *some* $i \in \{1, \dots, m\}$ *satisfies* $\|x - y_i\|_\infty < \varepsilon/\sqrt{n}$, *and* $m \le (C/\varepsilon)^{n-1}$ *for an absolute constant* C.

Hint: use rounding toward zero as above; the relevant scaled lattice points have Euclidean norms in $[1-\varepsilon, 1]$. *Surround them by cubes with disjoint interiors and side* $\varepsilon/\sqrt{n}$ *and compare their total volume with that of the resulting Euclidean shell. For* $n = 1$, *use the two points of* $\mathbb{S}^0$.

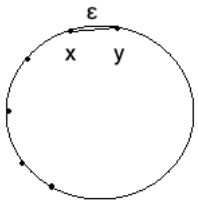
Remark 2.18. *The centers* y_i *need not lie on* $\mathbb{S}^{n-1}$. *In contrast, the usual Euclidean sphere net can be chosen with all centers on the sphere.*

We can also formulate the Euclidean and lattice net constructions in the following manner:

Theorem 2.19. *Let* $0 < \varepsilon < 1/2$.

1. *There exist points* $y_1, \dots, y_m \in \mathbb{S}^{n-1}$ *with* $m \le (C/\varepsilon)^n$ *such that every* $x \in \mathbb{S}^{n-1}$ *satisfies* $|x - y_i| < \varepsilon$ *for some* $i \in \{1, \dots, m\}$.

2. *There exist points* $y_1, \dots, y_m \in \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n$ *with* $m \le (C'/\varepsilon)^n$ *such that every* $x \in \mathbb{S}^{n-1}$ *satisfies* $\|x - y_i\|_\infty \le \varepsilon/\sqrt{n}$ *for some* i. *Moreover, the collection can be chosen to contain every vertex of every lattice cube of side* $\varepsilon/\sqrt{n}$ *that meets the sphere.*



Proof. For the first part, repeat the maximal packing argument of Theorem 2.8 with centers restricted to the sphere, using balls of radius $\varepsilon/4$ to obtain a strict ε-approximation.

For the second part, let $\delta = \varepsilon/\sqrt{n}$ and take all vertices of all cubes $\delta \prod_{j=1}^{n}[t_j, t_j + 1]$, $t_j \in \mathbb{Z}$, that meet $\mathbb{S}^{n-1}$. Each such vertex y is within Euclidean distance $\varepsilon < 1/2$ of a point on the sphere, so $1/2 < |y| < 3/2$. Also $y = \delta z$ for some $z \in \mathbb{Z}^n$ satisfying

$$\|z\|_1 \le \frac{\sqrt{n}|y|}{\delta} \le \frac{3n}{2\varepsilon}.$$

Theorem 2.14 bounds the number of these vertices by $(C'/\varepsilon)^n$. Every point of the sphere lies in one of the selected cubes, whose vertices are at ℓ_∞-distance at most δ from that point. □

Remark 2.20. *Note that $|a|/\sqrt{n} \le \|a\|_\infty \le |a|$. Thus the second part controls every coordinate on the finer scale $\varepsilon/\sqrt{n}$, while still giving Euclidean error at most ε and the same order of cardinality. Its centers, however, need not lie on the sphere.*

Remark 2.21. *In the second construction we keep* all vertices *of every lattice cube that meets the sphere. Some vertices may lie outside $\mathbf{B}_2^n$, but all lie in $\frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n$. Keeping all these vertices, rather than just one inward-rounded point from each cube, will allow us to use random rounding in the next application.*

2.7 An application: an efficient net for matrix multiplication

From now on, we will use the word *net* to mean a collection of points which has some good approximating properties for a given set (usually, the sphere). We will use notation $\mathcal{N}$ for a finite set of points, in place of just listing the points $\{y_1, \ldots, y_m\}$ as before, to save time. First, we outline the following

Theorem 2.22. *For every $0 < \varepsilon < 1$, there exists a set $\mathcal{N} \subseteq \mathbb{S}^{n-1}$ such that $\#\mathcal{N} \le (\frac{c}{\varepsilon})^n$, and for all $x \in \mathbb{S}^{n-1}$, there exists $y \in \mathcal{N}$ such that for any integer $N \ge 1$, and for any $N \times n$ matrix $\mathbf{A}$,*

$$\|\mathbf{A}(x - y)\| \le \|\mathbf{A}\|_{op} \cdot \varepsilon,$$

where we recall that the operator norm of a matrix was defined in (1).

Proof. By the usual net argument, for all $x \in \mathbb{S}^{n-1}$, there exists some $y \in \mathcal{N}$ such that $|x - y| < \varepsilon$. For any such matrix $\mathbf{A}$, we have $|\mathbf{A}(x - y)| \le \|\mathbf{A}\|_{op} \cdot |x - y| \le \|\mathbf{A}\|_{op} \cdot \varepsilon$. □

Next, we now explain a more precise way to approximate points on the sphere in order to compare the functional $|\mathbf{A}x|$ at those points. This net construction combines the ideas we already explained in Theorem 2.19 (part 2), with the idea of random rounding which we discussed in subsection 2.2. This net construction first appears in Klartag and Livshyts [19], and was later further developed in Livshyts [26].

Theorem 2.23. *Fix $n \in \mathbb{N}$ and $0 < \varepsilon < 1/2$. There exists $\mathcal{N} \subset \frac{3}{2}\mathbf{B}_2^n \backslash \frac{1}{2}\mathbf{B}_2^n$ such that $\#\mathcal{N} \leq (\frac{C}{\varepsilon})^n$, and so that for all $x \in \mathbb{S}^{n-1}$ and for all $N \in \mathbb{N}$, for any $N \times n$ matrix $\mathbf{A}$, there exists $y \in \mathcal{N}$ such that*

$$|\mathbf{A}(x-y)| \leq \varepsilon \cdot \frac{\|\mathbf{A}\|_{HS}}{\sqrt{n}},$$

where the Hilbert-Schmidt norm of a matrix was defined in Definition 1.2.

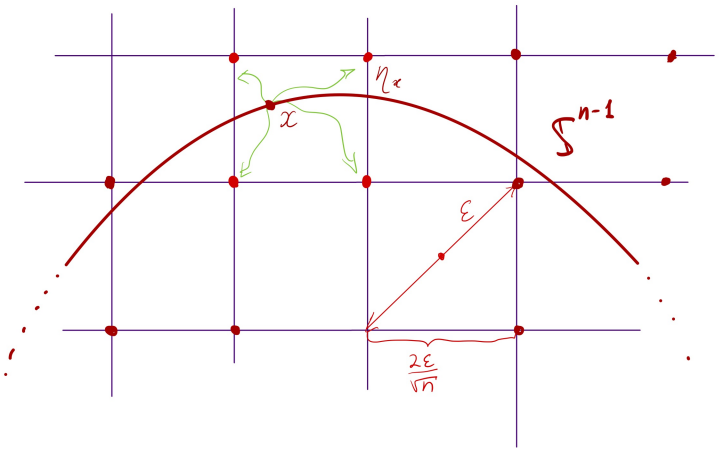
Remark 2.24. *As we noted before, $\frac{\|\mathbf{A}\|_{HS}}{\sqrt{n}} \cdot \varepsilon \leq \|\mathbf{A}\|_{op} \cdot \varepsilon$, and thus the net in Theorem 2.23 is more precise than the net in Theorem 2.22. Apart from some minor differences in the order of quantifiers (which we discuss below), Theorem 2.23 is stronger than Theorem 2.22, in the essential sense (but not literally).*

Proof of Theorem 2.23. Consider $\mathcal{N}$ to be the collection of lattice vertices constructed in the proof of Theorem 2.19, part 2. Then indeed $\mathcal{N} \subset \frac{3}{2}\mathbf{B}_2^n \backslash \frac{1}{2}\mathbf{B}_2^n$ such that $\#\mathcal{N} \leq (\frac{C}{\varepsilon})^n$, and for every $x \in \mathbb{S}^{n-1}$ one may find a lattice cube

$$Q = \frac{\varepsilon}{\sqrt{n}} \prod_{i=1}^{n} [t_i, t_i + 1],$$

for some integers t_i, such that $x \in Q$, and the vertices of Q all belong to the point set $\mathcal{N}$.

Consider the (scaled) random rounding η_x of $x \in \mathbb{S}^{n-1}$ to the vertices of the cube Q (as defined earlier).



Namely, η_x is a random vector which takes values in the vertices of Q, has independent coordinates, and $\mathbb{E}\,\eta_x = x$. Recall from (4), after rescaling by $\frac{\varepsilon}{\sqrt{n}}$, that

$$\mathbb{E}\langle x - \eta_x, \theta\rangle^2 \leq \frac{\varepsilon^2|\theta|^2}{4n} \leq \frac{\varepsilon^2|\theta|^2}{n}. \tag{8}$$

Recall that $|\mathbf{A}y|^2 = \sum_{i=1}^N \langle \mathbf{A}^T e_i, y\rangle^2$ (here $A^T e_i$ are the rows of A). Using the inequality (8) with $\theta = A^T e_i$, we get

$$\begin{aligned}
\mathbb{E}\,|\mathbf{A}(x-\eta_x)|^2 &= \mathbb{E}\sum_{i=1}^N \langle \mathbf{A}^T e_i, x-\eta_x\rangle^2 \\
&= \sum_{i=1}^N \mathbb{E}\langle \mathbf{A}^T e_i, x-\eta_x\rangle^2 \\
&\leq \frac{\varepsilon^2}{n}\sum_{i=1}^N |\mathbf{A}^T e_i|^2 \\
&= \|\mathbf{A}\|_{HS}^2 \cdot \frac{\varepsilon^2}{n}.
\end{aligned}$$

Since $\mathbb{E}\,|\mathbf{A}(x-\eta_x)|^2 \leq \|\mathbf{A}\|_{HS}^2 \cdot \frac{\varepsilon^2}{n}$, there exists a realization y of the random vector η_x such that

$$|\mathbf{A}(x-y)|^2 \leq \varepsilon^2 \cdot \frac{\|\mathbf{A}\|_{HS}^2}{n},$$

or equivalently,

$$|\mathbf{A}(x-y)| \leq \varepsilon \cdot \frac{\|\mathbf{A}\|_{HS}}{\sqrt{n}}.$$

Note that any realization of η_x is a vertex of Q and in particular, $y \in \mathcal{N}$. □

Remark 2.25. *Note that:*

- *For both Theorem 2.22 and Theorem 2.23, the net does not depend on the matrix* $\mathbf{A}$.
- *In Theorem 2.22, the approximation point* y *can be selected using* x *alone and then works for every matrix* $\mathbf{A}$. *In Theorem 2.23,* y *may depend on both* x *and* $\mathbf{A}$. *But we will see that in applications that we consider, this does not make a difference.*

Example 2.26. *Let us consider two concrete examples of matrices.*

- $\mathbf{A} = \mathrm{Id}_{\mathrm{n}}$. *Then* $\|\mathbf{A}\|_{op} = 1$, *and* $\|\mathbf{A}\|_{HS} = \sqrt{1+\cdots+1} = \sqrt{n}$. *Therefore,* $\frac{\|A\|_{HS}}{\sqrt{n}} = \|A\|_{op}$, *and therefore Theorems 2.22 and 2.23 provide the same precision.*

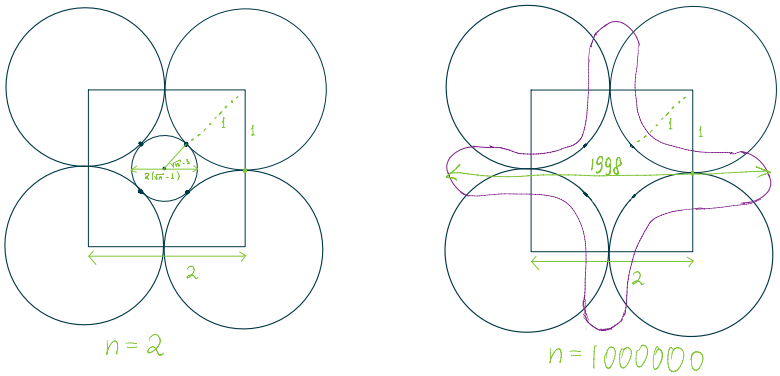
- $\mathbf{A} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$

  Then $\frac{1}{\sqrt{n}}\|\mathbf{A}\|_{HS} = \frac{1}{\sqrt{n}} \ll 1 = \|\mathbf{A}\|_{op}$*, and therefore,* ***Theorem 2.23 is more precise by a factor of*** $\sqrt{n}$ ***when approximating the first coordinate*** $x \mapsto x_1$ ***on the sphere, or, equivalently, any one-dimensional projection*** $\langle x, \xi \rangle$***!*** *For instance, this was crucial in [19], and the necessity to have a net with such a property motivated this construction.*

2.8 High-dimensional phenomenon: some themes

Weirdness of high dimension

Consider the cube B_∞^n and place a copy of B_2^n centered at each vertex of B_∞^n (see the Figure below). Note that the largest ball you can place at the center of B_∞^n with interior disjoint from the interiors of all the copies of B_2^n will have radius $\sqrt{n} - 1$. This might be somewhat surprising, since in many directions the ball extends out much further from the origin than the cube does, when n is large.



Central Limit Theorem and its geometric meaning

The following result is a very classical fact in Probability, see e.g. Durrett [7].

Theorem 2.27 (Central Limit Theorem)**.** *Let* $X_1, X_2, \ldots$ *be i.i.d. random variables with finite second moment. Let* $\mu = \mathbb{E}\, X_1$ *and* $\sigma^2 = \mathrm{Var}(X_1) > 0$*. Let* $S_n = \sum_{1 \le k \le n} X_k$*. Then,*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \to \mathcal{N}(0,1),$$

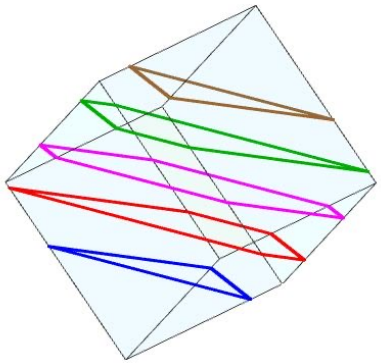
where the convergence is in distribution as $n \to \infty$.

To get a geometric interpretation, take $X_1, \dots, X_n$ to be independent and uniform on $[-1, 1]$. Then $X = (X_1, \dots, X_n)$ is uniform on the cube $\mathbf{B}_\infty^n = [-1, 1]^n$, and each coordinate has mean zero and variance $1/3$. For $\theta = (1/\sqrt{n}, \dots, 1/\sqrt{n})$, the theorem says that $\langle X, \theta \rangle$ converges in distribution to a normal random variable of variance $1/3$; equivalently, $\sqrt{3}\langle X, \theta \rangle$ converges to $\mathcal{N}(0, 1)$.

What is the geometric meaning of the density $f(t)$ of $\langle X, \theta \rangle$? Slicing the cube perpendicular to the unit vector θ gives

$$f(t) = \frac{|\mathbf{B}_\infty^n \cap \{x : \langle x, \theta \rangle = t\}|_{n-1}}{|\mathbf{B}_\infty^n|} = 2^{-n} |\mathbf{B}_\infty^n \cap (\theta^\perp + t\theta)|_{n-1}.$$

Thus $f(t)$ is the $(n-1)$-dimensional area of the corresponding section, normalized by the volume of the cube. Although individual sections of cubes can be hard to compute, the central limit theorem says that the volume proportions of slabs approach Gaussian probabilities. This is a precise sense in which their cross-sectional profile becomes simple in high dimension.



This phenomenon appears to stem from *independence*. However, this fact is true not just about cubes! For a uniform random point X in any high-dimensional convex body, after centering and applying a linear transformation so that the covariance is the identity, most directions θ have approximately Gaussian marginals $\langle X, \theta \rangle$. This is the content of the Central Limit Theorem for convex bodies from 2007 due to Klartag [18].

About the thin shell type concentration

Suppose that X is a random vector that is distributed uniformly over the unit ball $\mathbf{B}_2^n$. Then with high probability, its Euclidean norm is very close to 1 when n is large!

Indeed, for $n > 10$, note that

$$\begin{aligned} \mathbb{P}\left(|X| \geq 1 - \frac{10}{n}\right) = \frac{|\mathbf{B}_2^n \setminus (1 - 10/n)\mathbf{B}_2^n|}{|\mathbf{B}_2^n|} &= 1 - \left(1 - \frac{10}{n}\right)^n \\ &\geq 1 - e^{-10} \\ &\geq 0.99. \end{aligned}$$

3 Concentration Inequalities for sums of independent random variables

3.1 What is a concentration inequality?

In this section, we closely follow Vershynin [55]; see also the references therein. The letters $c, C > 0$ denote absolute constants, which may change from line to line. When a scale in the denominator of a tail bound is zero, the corresponding centered random variable is identically zero, and that case is understood separately. Recall the Law of Large Numbers (see Theorem 2.4.1 in Durrett [7]).

Theorem 3.1 (LLN)**.** *Let $(X_n)_{n\geq 1}$ be i.i.d. random variables with $\mathbb{E}\,|X_1| < \infty$. Then,*

$$\frac{X_1 + \ldots + X_N}{N} \xrightarrow{a.s.} \mathbb{E}\, X_1$$

as $N \to \infty$.

Definition 3.2. Let X be an integrable random variable. We say that X satisfies a **concentration inequality** if

$$\mathbb{P}\left(|X - \mathbb{E}\, X| > t\right) \leq \text{☺}(t).$$

We usually think of $\text{☺}(t)$ as a function that decays to 0 fast.

For sums of independent random variables, concentration inequalities give quantitative versions of the Law of Large Numbers. A classical example is Chebyshev's inequality:

Theorem 3.3 (Chebyshev)**.** *Let X be a random variable with $\mathbb{E}\,|X|^2 < \infty$ and $t > 0$. Then,*

$$\mathbb{P}\left(|X - \mathbb{E}\, X| \geq t\right) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proof. First, note that

$$\mathbb{E}\,|X - \mathbb{E}\, X|^2 = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 \leq \mathbb{E}\, X^2 < \infty.$$

Applying Markov's inequality to the random variable $|X - \mathbb{E}\, X|^2$, we see that

$$\begin{aligned}\mathbb{P}\left(|X - \mathbb{E}\, X| \geq t\right) &= \mathbb{P}\left(|X - \mathbb{E}\, X|^2 \geq t^2\right) \\ &\leq \frac{1}{t^2}\,\mathrm{Var}(X).\end{aligned}$$

□

3.2 Hoeffding's inequality

However, Chebyshev's inequality provides a weak bound when it comes to random variables which are sums of independent random variables with certain properties. Below, we present the first example of a much stronger result.

Theorem 3.4 (Hoeffding's Inequality)**.** *Let $X_1, \ldots, X_n$ be independent symmetric Bernoulli random variables (taking values 1 and -1 with probability 0.5 each). Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Then, for all $t \geq 0$, we have*

$$\mathbb{P}\left(\sum_{1 \leq i \leq n} a_i X_i \geq t\right) \leq e^{-\frac{t^2}{2|a|^2}}$$

Proof of Theorem 3.4. The cases $a = 0$ and $t = 0$ are immediate. Otherwise, by replacing a by $a/|a|$ and t by $t/|a|$, we may assume that $a \in \mathbb{S}^{n-1}$.

Then, for $\lambda > 0$, using the so-called Chernoff trick, we have

$$\begin{aligned}\mathbb{P}\left(\sum_i a_i X_i \geq t\right) &= \mathbb{P}\left(e^{\lambda \sum a_i X_i} \geq e^{\lambda t}\right) \\ &\leq e^{-\lambda t} \prod_{1 \leq i \leq n} \mathbb{E}\left(e^{\lambda a_i X_i}\right)\end{aligned}$$

Recall that $X_1, \ldots, X_n$ are symmetric Bernoulli random variables, i.e.,

$$X_i = \begin{cases} 1, & \text{w.p. } 1/2 \\ -1, & \text{w.p. } 1/2 \end{cases}$$

Hence, we have

$$\begin{aligned}\mathbb{E}\left(e^{\lambda a_i X_i}\right) &= \frac{1}{2} e^{\lambda a_i} + \frac{1}{2} e^{-\lambda a_i} \\ &= \cosh(\lambda a_i)\end{aligned}$$

Hence, for any $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\left(\sum_{1 \leq i \leq n} a_i X_i \geq t\right) &\leq e^{-\lambda t} \prod_{1 \leq i \leq n} \cosh(\lambda a_i) && (9) \\ &\leq e^{-\lambda t} \prod_{1 \leq i \leq n} e^{\lambda^2 a_i^2/2} && (\text{prove } \cosh(x) \leq e^{x^2/2} \text{ in HW}) \\ &= e^{-\lambda t} e^{\lambda^2/2 \sum_{1 \leq i \leq n} a_i^2} && (10) \\ &= e^{-\lambda t + \lambda^2/2}. && (|a| = 1)\end{aligned}$$

By picking the optimal $\lambda = t$, the upper bound becomes $e^{-t^2/2}$. □

Remark 3.5. *There is a nice geometric interpretation of this inequality. Let X be uniformly distributed on $\{-1,1\}^n$, $t>0$ and $a \in \mathbb{S}^{n-1}$.*

$$\mathbb{P}\left(\langle X, a\rangle \geq t\right) \leq e^{-t^2/2},$$

and therefore, the relative number of the points of the hypercube located in a half-space orthogonal to a and distance t from the origin behaves as (or, rather, better than) the Gaussian function.

Theorem 3.6 (Hoeffding's inequality for bounded random variables)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \in [m_i, M_i]$ almost surely, for some $m_i, M_i \in \mathbb{R}$. Then, for all $\beta > 0$, we have*

$$\mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}\, X_i)\right| \geq \beta\right) \leq 2e^{-\frac{2\beta^2}{\sum_i (M_i - m_i)^2}}.$$

Proof of Theorem 3.6. Homework.

Hint: first show that $\mathbb{E}\, e^{\lambda(X_i - \mathbb{E}\, X_i)} \leq e^{\lambda^2 (M_i - m_i)^2/8}$ for $\lambda \in \mathbb{R}$ (Hoeffding's lemma). Then use independence and Chernoff's trick. Finally,

$$\mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}\, X_i)\right| \geq \beta\right) = \mathbb{P}\left(\sum_i (X_i - \mathbb{E}\, X_i) \geq \beta\right) + \mathbb{P}\left(\sum_i (X_i - \mathbb{E}\, X_i) \leq -\beta\right),$$

so it remains to add the estimates for the two tails. □

Remark 3.7. *(Hoeffding vs. Chebyshev) Theorem 3.6 tells us that*

$$\mathbb{P}\left(\left|\sum_{i=1}^n (X_i - \mathbb{E}\, X_i)\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_i (M_i - m_i)^2}\right).$$

On the other hand, Chebyshev's inequality yields that

$$\mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}\, X_i)\right| \geq t\right) \leq \frac{\operatorname{Var}(\sum_i X_i)}{t^2} = \frac{\sum_i \operatorname{Var}(X_i)}{t^2}.$$

For fixed nonzero scales, the exponential bound decays faster than the bound proportional to t^{-2}. Thus Hoeffding's inequality gives a tighter bound for sufficiently large deviations, at the cost of the stronger assumptions on the summands.

Remark 3.8. *(Hoeffding vs. Central Limit Theorem) Let X_i be independent Bernoulli random variables with a common parameter $p \in (0,1)$, i.e.,*

$$X_i = \begin{cases} 1, & \text{w.p. } p \\ -1, & \text{w.p. } 1-p \end{cases}$$

Then $\mathbb{E}\, X_i = 2p-1$ *and* $\mathrm{Var}(X_i) = 4p(1-p)$*. According to the CLT,*

$$\frac{\sum_{i=1}^n X_i - n(2p-1)}{2\sqrt{np(1-p)}} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

For $p = 1/2$*, this is the normalization* $(X_1 + \cdots + X_n)/\sqrt{n}$*. For* $t > 0$*, observe that*

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{t} - \frac{1}{t^3}\right) e^{-\frac{t^2}{2}} \le \mathbb{P}(Z > t) \le \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-\frac{t^2}{2}}. \tag{11}$$

The lower bound is useful for $t > 1$ *and is left as homework. For the upper bound, note that*

$$\begin{aligned}
\mathbb{P}(Z > t) &= \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}\, \mathrm{d}s \\
&= \frac{1}{\sqrt{2\pi}} \int_t^\infty s \cdot \frac{1}{s} \cdot e^{-\frac{s^2}{2}}\, \mathrm{d}s \\
&\le \frac{1}{t} \frac{1}{\sqrt{2\pi}} \int_t^\infty s \cdot e^{-\frac{s^2}{2}}\, \mathrm{d}s \\
&= \frac{1}{t} \frac{1}{\sqrt{2\pi}} \int_{t^2/2}^\infty e^{-s}\, \mathrm{d}s \\
&= \frac{1}{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.
\end{aligned}$$

But we cannot directly deduce Hoeffding from CLT: although

$$\left| \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n(2p-1)}{2\sqrt{np(1-p)}} \ge t \right) - \mathbb{P}(Z \ge t) \right| \longrightarrow 0 \qquad (n \to \infty),$$

the CLT alone supplies neither an error bound for a fixed n *nor a useful relative error when the Gaussian tail is very small. Hoeffding's inequality gives a bound for every* n *and every threshold. In the symmetric case, its Gaussian decay agrees with the tail behaviour suggested by the CLT.*

Remark 3.9. *(HW) In fact, Hoeffding is not tight for non-symmetric Bernoulli random variables*

$$X_i = \begin{cases} 1 & \text{w.p. } p_i \\ -1 & \text{w.p. } 1 - p_i, \end{cases}$$

where $p_i \approx 0.99$*. Applying the Chernoff bounds below to the indicators* $(1 - X_i)/2$ *captures the small probability of a negative value and can give a better estimate for the centered sum.*

3.3 Chernoff's inequality

Theorem 3.10. *(Chernoff's Inequality) Let*

$$X_i = \begin{cases} 1, & w.p.\ p_i \\ 0, & w.p.\ 1-p_i \end{cases}$$

be independent Bernoulli random variables with parameters p_i. Denote $S_N = \sum_{i=1}^N X_i$ and

$$\mathbb{E}\, S_N = \sum_{i=1}^N \mathbb{E}\, X_i = \sum_{i=1}^N p_i =: \mu.$$

Then for all $t > \mu$,

$$\mathbb{P}(S_N > t) \le e^{-\mu} \left(\frac{e\mu}{t}\right)^t.$$

Proof. If $\mu = 0$, then $S_N = 0$ almost surely and the assertion is immediate. Assume $\mu > 0$. For any $\lambda > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sum X_i > t\right) = \mathbb{P}\left(e^{\lambda \sum X_i} > e^{\lambda t}\right) &\le e^{-\lambda t}\, \mathbb{E}\, e^{\lambda \sum X_i} \\
&= e^{-\lambda t} \prod_{i=1}^N \mathbb{E}\, e^{\lambda X_i} && \text{(by independence)} \\
&= e^{-\lambda t} \prod_{i=1}^N (p_i e^\lambda + 1 - p_i) \\
&= e^{-\lambda t} \prod_{i=1}^N (1 + p_i(e^\lambda - 1)) \\
&\le e^{-\lambda t} \prod_{i=1}^N e^{p_i(e^\lambda - 1)} && (1 + x \le e^x, \forall x \in \mathbb{R}) \\
&= e^{-\lambda t} e^{(e^\lambda - 1)\sum p_i} \\
&= e^{-\lambda t} e^{\mu(e^\lambda - 1)}
\end{aligned}$$

By choosing the optimal $\lambda = \log \frac{t}{\mu}$ (HW: show that this choice is indeed optimal), the upper bound becomes $e^{-\mu}(\frac{e\mu}{t})^t$. □

Corollary 3.11 (HW, "small deviation Chernoff")**.** *Suppose*

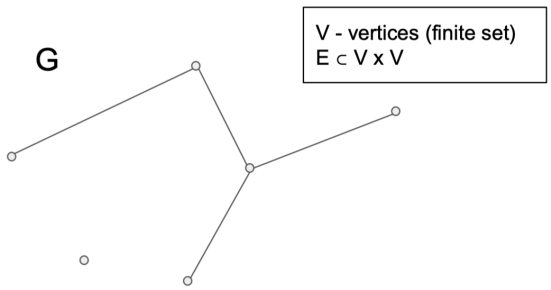
$$X_i = \begin{cases} 1, & w.p.\ p_i \\ 0, & w.p.\ 1-p_i \end{cases},$$

and $X_1, \cdots, X_N$ are independent. Let $S_N = \sum_{i=1}^N X_i$ and $\mathbb{E}\, S_N = \mu$. Then, for all $\delta \in [0,1]$,

$$\mathbb{P}\left(|S_N - \mu| \ge \delta\mu\right) \le 2e^{-c\mu\delta^2}.$$

3.4 Application to Random Graphs

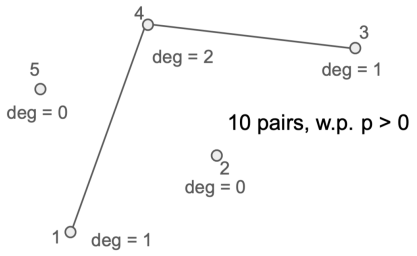
A simple undirected graph is a pair (V, E), where V (vertices) is a finite set and E (edges) is a collection of unordered pairs of distinct vertices. Schematically, if $\{v_i, v_j\} \in E$, then there is an edge between v_i and v_j.



A random graph is a graph selected in some random way.

Definition 3.12 (Erdös-Rényi Random Graph)**.** Given a positive integer n and $p \in [0, 1]$, a **random graph** $\mathbb{G}(n, p)$ satisfies the properties:

1. the set of vertices is $\{1, \ldots, n\}$.
2. For each unordered pair of distinct vertices, there is an edge with probability p.
3. The edge events are independent.



Recall that the degree of a vertex v, denoted by $\deg(v)$, is the number of edges incident to v.

Theorem 3.13 (The degree of each vertex of $G(n, p)$ is well concentrated around its mean)**.** *Let $n \geq 2$ and $G \sim \mathbb{G}(n, p)$. Assume that*

$$\mathbb{E}[\deg(v)] = d = (n-1)p \geq C \log n,$$

where $C > 0$ is a sufficiently large absolute constant. Then, for every $v \in V$,

$$\mathbb{P}\left(\deg(v) \in [0.9d, 1.1d]\right) \geq 0.9.$$

In fact, with probability at least 0.9, this inclusion holds for all vertices simultaneously.

Proof of Theorem 3.13. For $i \neq j$, let

$$X_{ij} = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1-p \end{cases}$$

be the indicator of an edge between the i-th and j-th vertices. For fixed i, these indicators are independent, and $\deg(i) = \sum_{j \neq i} X_{ij}$. By the small-deviation Chernoff bound (Corollary 3.11),

$$\mathbb{P}\left(|\deg(i) - d| \geq 0.1d\right) \leq 2e^{-cd}.$$

The union bound therefore gives

$$\mathbb{P}\left(\text{some } i \text{ satisfies } |\deg(i) - d| \geq 0.1d\right) \leq 2ne^{-cd} \leq 2n^{1-cC} \leq 0.1$$

when C is sufficiently large. No independence between the degrees is needed for the union bound. □

Remark 3.14. *Since $\mathbb{E}[\deg(v)] = p(n-1)$, the condition $d \geq C \log n$ is exactly $p \geq C \log n/(n-1)$. Up to an absolute constant, this is the familiar threshold $p \geq C' \log n/n$.*

3.5 Sub-Gaussian Random Variables

Next, we busy ourselves with the question:

Could we prove Hoeffding's inequality in a more general situation than bounded random variables? What sort of generality could we hope for?

Note that if the inequality

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| \geq t\right) \leq 2e^{-ct^2}$$

holds for all $a \in \mathbb{S}^{n-1}$ and $t \geq 0$, then by taking $a = e_i$, we see that necessarily

$$\mathbb{P}(|X_i| \geq t) \leq 2e^{-ct^2}. \tag{12}$$

We will show that (12) is also sufficient, up to changing the absolute constant, when $X_1, \ldots, X_n$ are independent and mean zero. Without centering, we obtain the corresponding bound for $\sum_i a_i(X_i - \mathbb{E}\, X_i)$. As a first step, we show the following (see Proposition 2.5.2 of [55]):

Proposition 3.15. *Let X be a random variable. Then, the following are equivalent. More precisely, if one property holds with a positive parameter K_i, the others hold with parameters $K_j \leq CK_i$, where C is an absolute constant:*

1. $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/K_1^2}$, *for all* $t \geq 0$ *and a fixed* $K_1 > 0$;
2. $||X||_p = (\mathbb{E}\,|X|^p)^{1/p} \leq K_2\sqrt{p}$ *for all* $p \geq 1$;
3. $\mathbb{E}\,e^{\lambda^2 X^2} \leq e^{K_3^2\lambda^2}$ *for all* $\lambda \in [-\frac{1}{K_3}, \frac{1}{K_3}]$;
4. $\mathbb{E}\,e^{\frac{X^2}{K_4^2}} \leq 2$.

   Moreover, if $\mathbb{E}\,X = 0$, *then conditions (1) - (4) are also equivalent to*

5. $\mathbb{E}\,e^{\lambda X} \leq e^{K_5^2\lambda^2}$ *for all* $\lambda \in \mathbb{R}$.

Proof. First we show (1) $\Rightarrow$ (2). By scaling, we may normalize the parameter in each assumed property to be 1. Suppose that $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2}$. We would like to show that in this case, $\mathbb{E}\,|X|^p \leq (C\sqrt{p})^p$. Indeed, by (3), and using a change of variables, we get

$$\begin{aligned}\mathbb{E}\,|X|^p &= \int_0^\infty \mathbb{P}(|X|^p > t)\ \mathrm{d}t \\ &= p\int_0^\infty \mathbb{P}(|X| > s)s^{p-1}\ \mathrm{d}s \\ &\leq 2p\int_0^\infty s^{p-1}e^{-s^2}\ \mathrm{d}s = p\Gamma\left(\frac{p}{2}\right).\end{aligned}$$

Here we used the tail assumption and the substitution $u = s^2$. Recall that

$$\Gamma(m) = \int_0^\infty t^{m-1}e^{-t}\ \mathrm{d}t \qquad (m > 0).$$

Stirling's estimate, together with boundedness on a fixed compact interval, gives $p\Gamma(p/2) \leq (C\sqrt{p})^p$ for every $p \geq 1$.

Next we show (2) $\Rightarrow$ (3). Suppose $(\mathbb{E}|X|^p)^{\frac{1}{p}} \leq \sqrt{p}$. By Taylor's formula, $e^a = \sum_{k=0}^\infty a^k/k!$ for every $a \geq 0$. Since the terms are nonnegative, we may exchange expectation and sum-

mation, and therefore

$$\begin{aligned}
\mathbb{E}e^{\lambda^2 X^2} &= \mathbb{E}\sum_{k=0}^{\infty}\frac{(\lambda^2 X^2)^k}{k!} = \sum_{k=0}^{\infty}\frac{\lambda^{2k}\mathbb{E}X^{2k}}{k!} \\
&\le 1 + \sum_{k=1}^{\infty}\frac{\lambda^{2k}(2k)^k}{k!} \\
&\le \sum_{k=0}^{\infty}(2e\lambda^2)^k && (k! \ge (k/e)^k) \\
&= \frac{1}{1-2e\lambda^2} \\
&\le e^{4e\lambda^2} \qquad \text{if } |\lambda| \le \frac{1}{2\sqrt{e}}.
\end{aligned}$$

In the last passage we used $(1-x)^{-1} \le e^{2x}$ for $0 \le x \le 1/2$ (homework). This proves (3) with $K_3 = 2\sqrt{e}$ in the normalized case.

It is straightforward to show (3) $\Rightarrow$ (4): take $\lambda = \sqrt{\log 2}/K_3$, so that $K_4 = K_3/\sqrt{\log 2}$ works.

To verify (4) $\Rightarrow$ (1), note that

$$\mathbb{P}(|X| \ge t) = \mathbb{P}(e^{X^2/K_4^2} \ge e^{t^2/K_4^2}) \le \mathbb{E}\, e^{X^2/K_4^2} e^{-t^2/K_4^2} \le 2e^{-t^2/K_4^2},$$

where the last inequality uses (4).

Now check (3) $\Rightarrow$ (5). Suppose $\mathbb{E}\,X = 0$ and $\mathbb{E}\,e^{\lambda^2 X^2} \le e^{\lambda^2}$ for $|\lambda| \le 1$. We use the inequality $e^x \le x + e^{x^2}$ for all $x \in \mathbb{R}$ (which we leave as homework).

$$\begin{aligned}
\mathbb{E}\,e^{\lambda X} &\le \mathbb{E}(\lambda X + e^{\lambda^2 X^2}) \\
&= \mathbb{E}\,e^{\lambda^2 X^2} && (\mathbb{E}\,X = 0) \\
&\le e^{\lambda^2} \qquad (|\lambda| \le 1).
\end{aligned}$$

Now assume $|\lambda| \ge 1$. Then we use the inequality $2\lambda x \le x^2 + \lambda^2, \forall \lambda, x \in \mathbb{R}$, which is also left as a homework.

$$\begin{aligned}
\mathbb{E}\,e^{\lambda X} &\le e^{\lambda^2/2}\,\mathbb{E}\,e^{X^2/2} \\
&\le e^{\lambda^2/2}\sqrt{e} \le e^{\lambda^2}. && (\text{apply (3)})
\end{aligned}$$

Lastly, we show (5) $\Rightarrow$ (1). Suppose $\mathbb{E}\,e^{\lambda X} \le e^{\lambda^2}$ for all $\lambda \in \mathbb{R}$. For $t > 0$ and $\lambda > 0$,

$$\mathbb{P}(X \ge t) = \mathbb{P}(e^{\lambda X} \ge e^{\lambda t}) \le e^{-\lambda t}\mathbb{E}e^{\lambda X} \le e^{-\lambda t}e^{\lambda^2}.$$

We plug the optimal value of $\lambda = \frac{t}{2}$, which gives $\mathbb{P}(X \geq t) \leq e^{-t^2/4}$. Applying the same argument to the lower tail, we get

$$\mathbb{P}(|X| \geq t) \leq \mathbb{P}(X \geq t) + \mathbb{P}(X \leq -t) \leq 2e^{-t^2/4}.$$

The case $t = 0$ is immediate. □

This brings us to a

Definition 3.16. (Sub-Gaussian random variables) If a random variable X satisfies any of properties (1)–(4) in Proposition 3.15, it is called a **sub-Gaussian** random variable. The sub-Gaussian norm of X, denoted $\|X\|_{\psi_2}$, is defined by

$$\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leq 2\}.$$

The sub-Gaussian norm is indeed a norm, which is left as homework.

Example 3.17 (Examples of sub-Gaussian variables)**.** *The following random variables are sub-Gaussian:*

- ***Gaussian****: $X \sim N(0,1)$ is sub-Gaussian; in fact, $\|X\|_{\psi_2} = \sqrt{8/3}$ with our normalization.*
- ***Bernoulli****: Let X be a symmetric Bernoulli random variable taking values ± 1. Since $|X| = 1$, it follows that X is a sub-Gaussian random variable with*

  $$\|X\|_{\psi_2} = \frac{1}{\sqrt{\ln 2}}.$$

- ***Bounded****: More generally, any bounded random variable X with $|X| \leq M$ almost surely, is sub-Gaussian with*

  $$\|X\|_{\psi_2} \leq C\|X\|_\infty$$

  where $C = \frac{1}{\sqrt{\ln 2}}$.

Next, we present

Example 3.18 (Examples of random variables that are not sub-Gaussian)**.** *The following random variables are not sub-Gaussian:*

- ***Exponential****: An exponential random variable is not sub-Gaussian. For example, the probability density function of an exponential random variable of mean 1 is*

  $$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ e^{-x} & \text{if } x \geq 0 \end{cases}.$$

  And the tail probability is given by:

  $$\mathbb{P}(X \geq t) = \int_t^\infty e^{-s}\, \mathrm{d}s = e^{-t} \qquad (t \geq 0).$$

  For every $c > 0$, this exceeds $2e^{-ct^2}$ for sufficiently large t, so the sub-Gaussian tail bound fails.

- ***Poisson:*** *Poisson random variables with positive parameter are not sub-Gaussian.*
- ***Cauchy:*** *Cauchy random variables are not sub-Gaussian.*

3.6 General Hoeffding's inequality and Khintchine's inequality

Recall the fact that a sum of independent normal random variables X_i is normal. More precisely, if $X_i \sim N(0, \sigma_i^2)$ are independent then

$$\sum_{i=1}^{N} X_i \sim N\left(0, \sum_{i=1}^{N} \sigma_i^2\right). \tag{13}$$

This property of the normal distribution extends to general sub-Gaussian distributions (see also Proposition 2.6.1 of [55]):

Proposition 3.19. *(Sums of independent sub-Gaussians) Let $X_1, \ldots, X_N$ be independent, mean zero, sub-Gaussian random variables. Then $\sum_{i=1}^{N} X_i$ is also a sub-Gaussian random variable, and*

$$\left\| \sum_{i=1}^{N} X_i \right\|_{\psi_2}^2 \leq C \sum_{i=1}^{N} ||X_i||_{\psi_2}^2.$$

Proof.

$$\begin{aligned} \mathbb{E} \exp\left(\lambda \sum_{i=1}^{N} X_i\right) &= \prod_{i=1}^{N} \mathbb{E} \exp(\lambda X_i) && \text{(by independence)} \\ &\leq \prod_{i=1}^{N} \exp\left(C^2 \|\lambda X_i\|_{\psi_2}^2\right) = \exp\left(\lambda^2 K^2\right), \end{aligned}$$

where $K^2 := C^2 \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2$. By equivalence of properties (4) and (5) in Proposition 3.15 and Definition 3.16, we see that the sum $\sum_{i=1}^{N} X_i$ is sub-Gaussian, and

$$\left\| \sum_{i=1}^{N} X_i \right\|_{\psi_2} \leq C_1 K.$$

□

Remark 3.20. *If $X_i \sim N(0, \sigma_i^2)$, then $\|X_i\|_{\psi_2}^2 = (8/3)\sigma_i^2$ with our definition. Indeed, $\mathbb{E}\, e^{X_i^2/t^2} = (1 - 2\sigma_i^2/t^2)^{-1/2}$ when $t^2 > 2\sigma_i^2$. Thus, for independent centered Gaussians, $\|\sum_i X_i\|_{\psi_2}^2 = \sum_i \|X_i\|_{\psi_2}^2$ exactly.*

Proposition 3.19 is, in fact, nothing but the generalized form of the Hoeffding inequality!

Corollary 3.21. *(General Hoeffding's inequality) Let $X_1, \dots, X_N$ be independent, mean zero, sub-Gaussian random variables, and $a = (a_1, \dots, a_N) \in \mathbb{R}^N$. Then, for every $t \geq 0$, we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^N a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{ct^2}{K^2|a|^2}\right),$$

where $K = \max_i \|X_i\|_{\psi_2}$.

Proof.

$$\mathbb{P}\left(\left|\sum_{i=1}^N a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{ct^2}{K^2|a|^2}\right)$$

is equivalent to (1) in Proposition 3.15. By Proposition 3.19,

$$\left\|\sum_{i=1}^N a_i X_i\right\|_{\psi_2}^2 \leq C\sum_{i=1}^N \|a_i X_i\|_{\psi_2}^2 = C\sum_{i=1}^N a_i^2 \|X_i\|_{\psi_2}^2 \leq CK^2|a|^2.$$

□

Remark 3.22. *Note that in some cases, a tighter version of Hoeffding's inequality, which is evident from the argument above, may be useful: if $X_1, \dots, X_N$ are independent, mean zero, sub-Gaussian random variables, and $a = (a_1, \dots, a_N) \in \mathbb{R}^N$, then, for every $t \geq 0$, we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^N a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{ct^2}{\sum_{i=1}^N a_i^2 \|X_i\|_{\psi_2}^2}\right).$$

As an immediate corollary of Proposition 3.19 and the equivalence of the sub-Gaussian properties (2) and (4) in Proposition 3.15, we get:

Theorem 3.23 (Khintchine's inequality)**.** *Let $X_1, \dots, X_N$ be independent sub-Gaussian random variables with zero means and unit variances, and let $a = (a_1, \dots, a_N) \in \mathbb{R}^N$. Then for every $p \in [2, \infty)$ we have*

$$\left(\sum_{i=1}^N a_i^2\right)^{1/2} \leq \left\|\sum_{i=1}^N a_i X_i\right\|_{L_p} \leq CK\sqrt{p}\left(\sum_{i=1}^N a_i^2\right)^{1/2}$$

where $K = \max_i \|X_i\|_{\psi_2}$ and C is an absolute constant.

The upper bound follows from the preceding results. For the lower bound, use $\|Y\|_{L_p} \geq \|Y\|_{L_2}$ for $p \geq 2$ and $\mathbb{E}(\sum_i a_i X_i)^2 = \sum_i a_i^2$.

Remark 3.24 (Centering)**.** *In many results above we impose the assumption $\mathbb{E}\, X = 0$. For a noncentered variable, we can instead apply the result to its centered version. Indeed, if $X_1, \dots, X_n$ are sub-Gaussian, set $Y_i = X_i - \mathbb{E}\, X_i$. It can be verified (see Lemma 2.6.8 of [55]) that the Y_i are also sub-Gaussian, and*

$$\|Y_i\|_{\psi_2} = \|X_i - \mathbb{E}\, X_i\|_{\psi_2} \leq \|X_i\|_{\psi_2} + \|\mathbb{E}\, X_i\|_{\psi_2} \leq 2\|X_i\|_{\psi_2}.$$

Here the last step follows from Jensen's inequality. The resulting concentration bounds concern deviations from the mean.

3.7 Sub-Exponential Random Variables

When independent random variables are not sub-Gaussian, we can no longer apply Hoeffding's inequality to bound the tail probability for their sum. However, there are important situations when this needs to be done, and is also (as we will see) possible. For example, for a standard Gaussian random vector $X \in \mathbb{R}^n$, that is, a random vector with i.i.d. coordinates $X_i \sim \mathcal{N}(0,1)$, consider the random variable $|X|$ – the length of the standard Gaussian random vector. We could study $|X|^2 = \sum_{i=1}^n X_i^2$, which is the sum of i.i.d. random variables, but unfortunately we cannot apply Hoeffding's inequality: X_i^2 are not sub-Gaussian. Indeed, the tails of X_i^2 decrease only exponentially fast:

$$\mathbb{P}(X_i^2 \geq t) = \mathbb{P}(|X_i| \geq \sqrt{t}) \leq 2e^{-t/2} \qquad (t \geq 0).$$

More precisely, (11) shows that this tail is comparable to $t^{-1/2}e^{-t/2}$ for large t, so it cannot satisfy a sub-Gaussian bound in t. In this section, we are going to discuss such "sub-exponential" random variables, and study their properties, which, as it turns out, are also very nice.

Definition 3.25 (Sub-exponential random variable)**.** A random variable X is called **sub-exponential** if there exists $K > 0$ such that

$$\mathbb{P}(|X| > t) \leq 2e^{-\frac{t}{K}}$$

for all $t > 0$.

Example 3.26 (Examples of sub-exponential random variables)**.**

- *Let X be an exponential random variable, i.e., X has probability density function*

$$f_X(t) = \begin{cases} 0, & t < 0 \\ e^{-t}, & t \geq 0 \end{cases}$$

  Since $\mathbb{P}(X > s) = e^{-s}$, X is sub-exponential.

- *If $X \sim \mathcal{N}(0,1)$, then X^2 is sub-exponential.*

Proposition 3.27 (Equivalent definitions of sub-exponential random variables)**.** *For a random variable X, the following are equivalent. If one property holds with a positive parameter K_i, the others hold with parameters at most an absolute constant times K_i:*

(a) $\mathbb{P}(|X| \geq t) \leq 2e^{-\frac{t}{K_1}}$, *for all* $t > 0$;

(b) $(\mathbb{E}|X|^p)^{\frac{1}{p}} \leq K_2 p$ *for all* $p \geq 1$;

(c) $\mathbb{E}e^{\lambda|X|} \leq e^{K_3\lambda}$, *for all* $\lambda \in [0, \frac{1}{K_3}]$;

(d) $\mathbb{E}e^{|X|/K_4} \leq 2$*;*

If we further assume $\mathbb{E}X = 0$*, then properties (a)-(d) are equivalent to*

(e) $\mathbb{E}(e^{\lambda X}) \leq e^{K_5^2 \lambda^2}$ *if* $|\lambda| \leq \frac{1}{K_5}$.

The proof of the proposition is left as an exercise.

Definition 3.28. (ψ_1-norm) We define the sub-exponential norm by

$$\|X\|_{\psi_1} := \inf\{K > 0 : \mathbb{E}\, e^{|X|/K} \leq 2\}.$$

Checking that this is a norm is left as an exercise. The smallest admissible tail parameter in property (a) is comparable to this norm, but need not be equal to it.

Note that a sub-Gaussian random variable is always sub-exponential. But more is true:

Lemma 3.29. *A random variable* X *is sub-Gaussian if and only if* X^2 *is sub-exponential, and in fact* $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$.

Proof. Suppose that X is sub-Gaussian. Then, for all $t > 0$, $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/K^2}$ for some constant $K > 0$. Therefore,

$$\mathbb{P}(X^2 \geq t) = \mathbb{P}(|X| \geq \sqrt{t}) \leq 2e^{-\frac{t}{K^2}}.$$

Therefore, X^2 is sub-exponential. Reversing the calculation above establishes the reverse implication. For the exact norm identity, use the definitions: $\mathbb{E}\, e^{|X^2|/K} \leq 2$ if and only if $\mathbb{E}\, e^{X^2/(\sqrt{K})^2} \leq 2$. □

More generally,

Lemma 3.30. *Suppose that* X, Y *are sub-Gaussian random variables. Then* XY *is sub-exponential, and*

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \cdot \|Y\|_{\psi_2} \tag{14}$$

Proof. If either norm is zero, the assertion is immediate. Otherwise, by homogeneity we may divide X and Y by their respective norms and assume $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. By the definition of the sub-Gaussian norm, we then have $\mathbb{E}\, e^{X^2} \leq 2$ and $\mathbb{E}\, e^{Y^2} \leq 2$. We claim that

$$\mathbb{E}(e^{|XY|}) \leq 2.$$

Indeed, recall Young's inequality

$$\frac{a^2 + b^2}{2} \geq |ab|$$

for all real numbers a, b, as follows from the fact that $(a-b)^2 \geq 0$ and $(a+b)^2 \geq 0$. Using this inequality twice, we get:

$$\begin{aligned}
\mathbb{E}(e^{|XY|}) &\leq \mathbb{E}\, e^{\frac{X^2+Y^2}{2}} && (\text{ Young's inequality}) \\
&= \mathbb{E}\left(e^{\frac{X^2}{2}} \cdot e^{\frac{Y^2}{2}}\right) \\
&\leq \frac{1}{2} \cdot \mathbb{E}(e^{X^2} + e^{Y^2}) && (\text{ Young's inequality}) \\
&= \frac{1}{2}\left(\mathbb{E}\, e^{X^2} + \mathbb{E}\, e^{Y^2}\right) \\
&\leq 2.
\end{aligned}$$

□

Remark 3.31. *The constant* 1 *in Lemma 3.30 uses the matching exponential-moment definitions of* $\|\cdot\|_{\psi_1}$ *and* $\|\cdot\|_{\psi_2}$*: property (4) in Proposition 3.15 and property (d) in Proposition 3.27. Independence of* X *and* Y *is not needed.*

Using other equivalent definitions gives an analogous estimate up to an absolute constant. This is left as homework.

Remark 3.32. *Analogously to the centering trick of sub-Gaussians, if* X *is sub-exponential, then* $X - \mathbb{E}X$ *is also sub-exponential, and*

$$\|X - \mathbb{E}(X)\|_{\psi_1} \leq C \cdot \|X\|_{\psi_1}$$

for some absolute constant C*. (homework)*

3.8 Bernstein's inequality

The Bernstein inequality for sub-exponential distributions provides a bound on the probability that the sum of independent sub-exponential random variables deviates from its expected value. One may draw a parallel between Bernstein's inequality and Hoeffding's inequality, both in terms of the statement and the proof method. Historically, Bernstein's work dates to the 1920s, Chernoff's paper appeared in 1952, and Hoeffding's inequality was published in 1963.

Theorem 3.33 (Bernstein's inequality)**.** *Let* $X = (X_1, X_2, \cdots, X_n) \in \mathbb{R}^n$ *be a random vector, where the* X_i *are independent sub-exponential random variables with* $\mathbb{E}\, X_i = 0$*. Then, for all* $t \geq 0$*,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2\exp\left(-c \cdot \min\left(\frac{t^2}{\sum_{i=1}^n \|X_i\|_{\psi_1}^2}, \frac{t}{\max_i \|X_i\|_{\psi_1}}\right)\right).$$

Here $c > 0$ *is an absolute constant.*

Proof. Let $K = \max_{1\le i\le n} \|X_i\|_{\psi_1}$. The cases $K = 0$ and $t = 0$ are immediate. For $t > 0$ and $\lambda > 0$, Markov's inequality gives

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) &= \mathbb{P}\left(\exp\left[\lambda \sum_{i=1}^n X_i\right] \ge e^{\lambda t}\right) \\ &\le e^{-\lambda t}\cdot \mathbb{E}\exp\left(\lambda\sum_{i=1}^n X_i\right) \\ &= e^{-\lambda t}\cdot \prod_{i=1}^n \mathbb{E}\, e^{\lambda X_i}.\end{aligned}$$

Since the X_i are centered, property (e) in Proposition 3.27 implies that

$$\mathbb{E}\, e^{\lambda X_i} \le e^{C_0\lambda^2\|X_i\|_{\psi_1}^2}$$

for all λ satisfying $|\lambda| \le c_0/K$, where $c_0, C_0 > 0$ are absolute constants. Thus, for $0 < \lambda \le c_0/K$,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) &\le e^{-\lambda t}\cdot \prod_{i=1}^n \mathbb{E}\, e^{\lambda X_i} \\ &\le e^{-\lambda t}\cdot \prod_{i=1}^n e^{C_0\lambda^2\|X_i\|_{\psi_1}^2} \\ &= e^{-\lambda t + C_0\lambda^2\sigma^2},\end{aligned}$$

where $\sigma^2 = \sum_{i=1}^n \|X_i\|_{\psi_1}^2$. To minimize the exponent under the constraint $0 < \lambda \le c_0/K$, we take $\lambda = \min\{t/(2C_0\sigma^2), c_0/K\}$. This choice satisfies $C_0\lambda^2\sigma^2 \le \lambda t/2$, so

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) \le \exp\left(-\min\left\{\frac{t^2}{4C_0\sigma^2}, \frac{c_0 t}{2K}\right\}\right).$$

By a similar reasoning, we see that

$$\mathbb{P}\left(\sum_{i=1}^n X_i \le -t\right) \le \exp\left(-\min\left\{\frac{t^2}{4C_0\sigma^2}, \frac{c_0 t}{2K}\right\}\right).$$

The theorem then follows from combining the two bounds above. □

Applying Bernstein's inequality to $a_i X_i$, we obtain the following inequality:

Corollary 3.34. *Let $X = (X_1, X_2, \cdots, X_n) \in \mathbb{R}^n$ be a random vector, where the X_i are independent and sub-exponential, $\mathbb{E}X_i = 0$, and $a \in \mathbb{R}^n$. Let $K = \max_i \|X_i\|_{\psi_1}$. Then for all $t \ge 0$,*

$$\mathbb{P}(|\langle X, a\rangle| \ge t) \le 2e^{-c\cdot\min\left\{\frac{t^2}{K^2|a|^2}, \frac{t}{K\|a\|_\infty}\right\}}$$

If $a = (\frac{1}{n}, \dots, \frac{1}{n})$, then we obtain the following inequality:

Corollary 3.35. *Under the assumptions of the preceding corollary,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2\exp\left(-cn \cdot \min\left\{\frac{t^2}{K^2}, \frac{t}{K}\right\}\right), \qquad t \geq 0.$$

If the X_i are bounded by a universal constant, then the bound in Bernstein's inequality can be further strengthened. For example, see Theorem 2.8.4 in [55].

3.9 Concentration of the norm of a random vector with independent sub-Gaussian coordinates

A model application of Bernstein's inequality is the following inequality:

Theorem 3.36 (Concentration of the norm of a random vector with independent sub–Gaussian coordinates)**.** *Let $X = (X_1, X_2, \dots, X_n)$ be a random vector, where the X_i are independent. Suppose that $\|X_i\|_{\psi_2} \leq K$ and $\mathbb{E}\, X_i^2 = 1$ for every i. Then,*

$$\big\| \, |X| - \sqrt{n} \, \big\|_{\psi_2} \leq CK^2,$$

where C is an absolute constant.

Remark 3.37. *Equivalently, the theorem states that*

$$\mathbb{P}\left(\big||X| - \sqrt{n}\big| \geq t\right) \leq 2e^{-ct^2/K^4}, \qquad t \geq 0,$$

i.e. X belongs, with high probability, to a thin shell around the sphere of radius $\sqrt{n}$. This result holds, and is interesting already when X is Gaussian.

Proof. Since each X_i is sub-Gaussian, we know that X_i^2 is sub-exponential and so is $X_i^2 - 1$. Furthermore,

$$\begin{aligned} \big\|X_i^2 - 1\big\|_{\psi_1} &\leq C\,\big\|X_i^2\big\|_{\psi_1} \\ &= C\,\|X_i\|_{\psi_2}^2 \\ &\leq CK^2. \end{aligned}$$

Since $\mathbb{E}\, X_i^2 = 1$, Jensen's inequality gives $K \geq 1/\sqrt{\log 2} > 1$. Applying Bernstein's inequality (Theorem 3.33) to $\frac{1}{n}\,|X|^2 - 1$, we see that for any $s > 0$

$$\mathbb{P}\left(\left|\frac{1}{n}|X|^2 - 1\right| \geq s\right) \leq 2e^{-\frac{cn}{K^4}\min\{s, s^2\}},$$

where we used that $K \geq 1$. Note that for all $z \geq 0$ and $\delta > 0$, $|z-1| \geq \delta$ implies $|z^2 - 1| \geq \max(\delta^2, \delta)$. Therefore,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}|X| - 1\right| \geq \delta\right) &\leq \mathbb{P}\left(\left|\frac{1}{n}|X|^2 - 1\right| \geq \max(\delta^2, \delta)\right) \\ &\leq 2e^{-\frac{cn}{K^4}\delta^2}\end{aligned}$$

Finally, taking $t = \delta\sqrt{n}$ yields

$$\mathbb{P}\left(\left||X| - \sqrt{n}\right| \geq t\right) \leq 2e^{-\frac{ct^2}{K^4}}.$$

□

3.10 Sub-Gaussian random vectors

A random vector is called **sub-Gaussian** if each of its one-dimensional projections is sub-Gaussian:

Definition 3.38. Let $X \in \mathbb{R}^n$ be a random vector. Then, we say that X is **sub-Gaussian** if for all $\theta \in \mathbb{S}^{n-1}$, the random variable $\langle X, \theta\rangle$ is sub-Gaussian. The associated sub-Gaussian norm $\|\cdot\|_{\psi_2}$ is defined by

$$\|X\|_{\psi_2} := \sup_{\theta \in \mathbb{S}^{n-1}} \|\langle X, \theta\rangle\|_{\psi_2}.$$

We will see that sub-Gaussian random vectors form a rich class which includes many examples. First, consider the standard Gaussian random vector $X \sim \mathcal{N}(0, \mathrm{Id})$ on $\mathbb{R}^n$: its coordinates are independent standard normal random variables. For every $\theta \in \mathbb{S}^{n-1}$, $\langle X, \theta\rangle \sim \mathcal{N}(0,1)$ by (13), so X is sub-Gaussian.

Proposition 3.39 (An example of a sub-Gaussian random vector)**.** *Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ have independent coordinates with $\mathbb{E}\, X_i = 0$ and $\|X_i\|_{\psi_2} \leq K$. Then X is a sub-Gaussian random vector and $\|X\|_{\psi_2} \leq CK$.*

Proof. By Proposition 3.19, for any $\theta \in \mathbb{S}^{n-1}$,

$$\|\langle X, \theta\rangle\|_{\psi_2}^2 = \left\|\sum_{i=1}^n \theta_i X_i\right\|_{\psi_2}^2 \leq C\sum_{i=1}^n |\theta_i|^2 \|X_i\|_{\psi_2}^2 \leq C \max_{1\leq i\leq n} \|X_i\|_{\psi_2}^2 \leq CK^2.$$

This implies the Proposition. □

Therefore, the random vector uniformly distributed on $\{-1, 1\}^n$ is sub-Gaussian with an absolute constant. The same holds for the uniform random vector in $[-1, 1]^n$. More generally, independent centered coordinates bounded by a common constant give a dimension-independent sub-Gaussian bound. A noncentered bounded random vector is still sub-Gaussian, but its norm also depends on its mean.

Remark 3.40. *What if we do not assume independence? Let $X = (X_1, \dots, X_n)$ have coordinates satisfying $\|X_i\|_{\psi_2} \le K$, possibly with dependencies. Since $\|\cdot\|_{\psi_2}$ is a norm, for every $\theta \in \mathbb{S}^{n-1}$ we have*

$$\|\langle X, \theta\rangle\|_{\psi_2} \le \sum_{i=1}^n |\theta_i| \cdot \|X_i\|_{\psi_2} \le K \cdot \sum_{i=1}^n |\theta_i| \le \sqrt{n}K.$$

In the last passage we used the fact that for any $\theta \in \mathbb{S}^{n-1}$,

$$\sum_{i=1}^n |\theta_i| \le \sqrt{n}\sqrt{\sum_{i=1}^n \theta_i^2} = \sqrt{n},$$

and in fact the equality in the above is attained for $\theta = (1/\sqrt{n}, ..., 1/\sqrt{n})$. Therefore,

$$\mathbb{P}(|\langle X, \theta\rangle| \ge t) \le 2e^{-\frac{ct^2}{nK^2}},$$

or in other words, $\|X\|_{\psi_2} \le \sqrt{n}K$. The loss of $\sqrt{n}$ can be necessary: for $X = (a, \dots, a)$ with a nonzero sub-Gaussian random variable a, the choice $\theta = (1, \dots, 1)/\sqrt{n}$ gives $\|\langle X, \theta\rangle\|_{\psi_2} = \sqrt{n}\|a\|_{\psi_2}$. This loss can be problematic for applications.

In light of Remark 3.40 one may wonder if there is a sub-Gaussian random vector with the sub-Gaussian constant which does not depend on the dimension but with dependencies among the coordinates. Below we show one such example:

Proposition 3.41. *The uniform distribution on the sphere $X \sim \mathrm{Unif}(\sqrt{n}\,\mathbb{S}^{n-1})$ is sub-Gaussian and $\|X\|_{\psi_2} \le C$ for an absolute constant $C > 0$.*

Remark 3.42. *The normalization $\sqrt{n}\mathbb{S}^{n-1}$ is natural because this means that $\mathbb{E}\, X_i^2 = 1$ for all i, so the Propositions 3.39 and 3.41 compare in a natural way.*

Proof. Let $g \sim \mathcal{N}(0, \mathrm{Id})$ in $\mathbb{R}^n$. By rotational invariance, $g/|g|$ is uniform on $\mathbb{S}^{n-1}$ (homework), so we may represent $X = \sqrt{n}\, g/|g|$. Fix $\theta \in \mathbb{S}^{n-1}$ and write $Z = \langle g, \theta\rangle$, a standard normal random variable. For $0 \le t \le \sqrt{n}$, split according to whether $|g| \ge \sqrt{n}/2$:

$$\begin{aligned}
\mathbb{P}\left(|\langle X, \theta\rangle| \ge t\right) &= \mathbb{P}\left(\frac{\sqrt{n}\,|Z|}{|g|} \ge t\right) \\
&\le \mathbb{P}(|Z| \ge t/2) + \mathbb{P}(|g| < \sqrt{n}/2) \\
&\le 2e^{-t^2/8} + 2e^{-cn} \\
&\le 4e^{-c't^2}.
\end{aligned}$$

Here we used the Gaussian tail bound and Theorem 3.36. No independence between Z and $|g|$ is required. For $t > \sqrt{n}$, the probability is zero because $|X| = \sqrt{n}$. Combining the displayed estimate with the trivial upper bound 1, and decreasing the absolute constant in the exponent, gives

$$\mathbb{P}\left(|\langle X, \theta\rangle| \ge t\right) \le 2e^{-c''t^2} \qquad (t \ge 0).$$

The equivalence of the sub-Gaussian definitions now gives $\|X\|_{\psi_2} \le C$. □

3.11 Grothendieck's inequality

We now apply our results about sub-Gaussian random vectors to deduce the following fact from linear algebra, which is useful in semi-definite programming (see Vershynin [55]).

Theorem 3.43 (Grothendieck's inequality)**.** *Suppose $A = (a_{ij})$ is an $m \times n$ matrix with real entries such that, for every $x \in \{-1, 1\}^m$ and $y \in \{-1, 1\}^n$,*

$$\left| \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j \right| \le 1.$$

Then, for unit vectors $u_1, \dots, u_m, v_1, \dots, v_n$ in any real Hilbert space,

$$\left| \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \langle u_i, v_j \rangle \right| \le K,$$

where $K > 0$ is an absolute constant.

Proof. First, the assumption can be equivalently stated as

$$\left| \sum_{i,j} a_{ij} x_i y_j \right| \le \max_i |x_i| \max_j |y_j|$$

for any real numbers $x_1, \dots, x_m, y_1, \dots, y_n$. Indeed, after scaling, a bilinear form on the product of two cubes attains its maximum absolute value at vertices. The corresponding homogeneous form of the conclusion is

$$\left| \sum_{i,j} a_{ij} \langle u_i, v_j \rangle \right| \le K \max_i \|u_i\| \max_j \|v_j\|.$$

If we allowed the constant to depend on A, we could take $K = \sum_{i,j} |a_{ij}|$. Fix A and let $K = K(A)$ be the smallest constant for which this homogeneous inequality holds for all vector families in any real Hilbert space. Thus K is already known to be finite. Our goal is to show $K \le 288$, following [55].

Any family of $m+n$ vectors spans a space of dimension at most $m+n$. We can therefore work in $\mathbb{R}^{m+n}$; by compactness, choose vectors of norms at most 1 such that

$$\sum_{i,j} a_{ij} \langle u_i, v_j \rangle = K.$$

We may change the signs of all the u_i to make the sum nonnegative.

Consider $U_i = \langle g, u_i \rangle$ and $V_j = \langle g, v_j \rangle$, where g is a standard Gaussian vector in $\mathbb{R}^{m+n}$. Then $\mathbb{E}(U_i V_j) = \langle u_i, v_j \rangle$ (homework), so

$$K = \sum_{i,j} a_{ij} \, \mathbb{E}(U_i V_j) = \mathbb{E} \sum_{i,j} a_{ij} U_i V_j.$$

Also $\mathbb{E}\,U_i^2 \le 1$ and $\mathbb{E}\,V_j^2 \le 1$.

If $|U_i| \le R$ and $|V_j| \le R$ almost surely, the scalar inequality would give

$$\left| \sum_{i,j} a_{ij} U_i V_j \right| \le R^2$$

pointwise, and hence $K \le R^2$. We approximate this situation by truncation.

Write $U_i = U_i^+ + U_i^-$, where

$$U_i^+ = U_i \mathbf{1}_{\{|U_i| \le R\}}, \qquad U_i^- = U_i \mathbf{1}_{\{|U_i| > R\}},$$

and similarly decompose $V_j = V_j^+ + V_j^-$. The superscripts here indicate the bounded and tail parts, not positive and negative parts. Then

$$\begin{aligned} K = \; & \mathbb{E} \sum_{i,j} a_{ij} U_i^+ V_j^+ + \mathbb{E} \sum_{i,j} a_{ij} U_i^+ V_j^- \\ & + \mathbb{E} \sum_{i,j} a_{ij} U_i^- V_j^+ + \mathbb{E} \sum_{i,j} a_{ij} U_i^- V_j^-. \end{aligned}$$

The first term is at most R^2 by the scalar inequality. For the other three terms, note that

$$\mathbb{E}\,|U_i^-|^2 \le \frac{\mathbb{E}\,U_i^4}{R^2} \le \frac{3}{R^2} \le \frac{4}{R^2},$$

and similarly for V_j^-. Moreover, every truncated part has L_2 norm at most 1.

Now use the defining inequality for $K(A)$ on the Hilbert space L_2, whose scalar product is $\langle Y, Z \rangle_{L_2} = \mathbb{E}(YZ)$. For example,

$$\begin{aligned} \left| \mathbb{E} \sum_{i,j} a_{ij} U_i^- V_j^+ \right| &= \left| \sum_{i,j} a_{ij} \langle U_i^-, V_j^+ \rangle_{L_2} \right| \\ &\le K \max_i \|U_i^-\|_{L_2} \max_j \|V_j^+\|_{L_2} \le \frac{2K}{R}. \end{aligned}$$

The other two terms containing a tail part obey the same bound. This does not assume the theorem we are proving: for the fixed matrix A, the finite constant $K(A)$ was defined to work in every Hilbert space; only its uniform bound is still to be proved.

Putting everything together gives

$$K \le R^2 + \frac{6K}{R}.$$

Choosing $R = 12$ and rearranging yields $K \le 288$, independently of A, m and n. □

4 Random Matrices

Definition 4.1. A random $N \times n$ matrix $A \in \mathbb{R}^{N \times n}$ is a matrix drawn in some random way. We shall use notation

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{Nn} \end{bmatrix} \tag{15}$$

where the entries a_{ij} are random variables.

Example 4.2. *Here are some examples of random matrices:*

- *When $a_{ij} \sim N(0,1)$ and are independent, A is sometimes called a Gaussian random matrix;*
- *More generally, we may consider a_{ij} to be independent random variables selected according to some distributions;*
- *We may also consider a symmetric random matrix, by selecting the entries on and above the diagonal independently and imposing the rule $a_{ij} = a_{ji}$;*
- *We may also select one random variable a and fill each entry with it;*
- *We may select some entries independently at random, while other entries would have fixed values (for example zeroes);*
- *Another example is a random rotation, selected according to Haar probability measure on the rotation group.*

We think about the random $N \times n$ matrix as a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^N$. We will study various properties of random matrices assuming that N and n are very large (but not reaching the infinity limit). This is what is informally called *non-asymptotic random matrix theory*. We will be using the methods of high-dimensional probability, one of many approaches to the study of random matrices. We will leave a lot of the relevant methods and questions and theories beyond the scope of this course.

For convenience we will assume throughout that $N \geq n$ since most of the properties of a matrix A which we study can be easily transferred to properties of A^T.

What properties of random matrices can be studied? One example is singular values. Recall that the singular values of a matrix $A \in \mathbb{R}^{N \times n}$ are given by $\sigma_i(A) = \sqrt{\lambda_i(A^T A)}$, where λ_i are the eigenvalues, and we suppose that $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$. Recall that

$$\begin{aligned} \sigma_1(A) &= \|A\| = \sup_{x \in \mathbb{S}^{n-1}} |Ax|; \\ \sigma_n(A) &= \inf_{x \in \mathbb{S}^{n-1}} |Ax|; \\ \sigma_i(A) &= \sup_{\dim(E)=i} \inf_{x \in \mathbb{S}(E)} |Ax|, \end{aligned} \tag{16}$$

where $\mathbb{S}(E)$ denotes the unit sphere in the subspace E. See Vershynin [55] for the details.

Definition 4.3. For a matrix A of full column rank, its condition number is $\kappa(A) = \frac{\sigma_1(A)}{\sigma_n(A)}$. We set $\kappa(A) = \infty$ when A does not have full column rank.

The condition number measures the ratio between the largest and smallest stretch factors. It is unchanged by multiplying A by a nonzero scalar; in particular, $\kappa(A) = 1$ precisely when A is a positive scalar multiple of an isometric embedding. The parameter $\kappa(A)$ is directly involved in the speed of various algorithms for solving systems of linear equations, and it is important in many applications to know that a certain random matrix model has small enough condition number with high probability.

4.1 Norm of a sub-Gaussian random matrix

We start by studying the norm of a class of random matrices.

Theorem 4.4 (Norm of matrices with independent mean zero sub-Gaussian entries)**.** *Let A be an $N \times n$ random matrix whose entries a_{ij} are independent, mean zero, sub-Gaussian random variables. Suppose $N \geq n$. Then, for any $t > 0$ we have*

$$\|A\| \leq CK(\sqrt{N} + t) \tag{17}$$

with probability at least $1 - 2e^{-t^2}$. Here $K = \max_{i,j} \|a_{ij}\|_{\psi_2}$ and $C > 0$ is some constant.

Proof. If $K = 0$, the assertion is immediate. Otherwise, replace A by A/K; we may assume $K = 1$ until the last step. By Theorem 5, with $\varepsilon = \frac{1}{4}$, we can find an ε-net $\mathcal{N} \subset \mathbb{S}^{n-1}$ and an ε-net $\mathcal{M} \subset \mathbb{S}^{N-1}$ where

$$\#\mathcal{N} \leq \left(\frac{2+\varepsilon}{\varepsilon}\right)^n, \ \#\mathcal{M} \leq \left(\frac{2+\varepsilon}{\varepsilon}\right)^N. \tag{18}$$

Note that

$$\sup_{x \in \mathbb{S}^{n-1}, y \in \mathbb{S}^{N-1}} \langle Ax, y\rangle \geq \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y\rangle,$$

since the supremum over a larger set is larger than the supremum over a smaller set. However,

$$\sup_{x \in \mathbb{S}^{n-1}, y \in \mathbb{S}^{N-1}} \langle Ax, y\rangle \cdot (1 - 2\varepsilon) \leq \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y\rangle. \tag{19}$$

Indeed, suppose $x \in \mathbb{S}^{n-1}, \tilde{x} \in \mathcal{N}$ so that $|x - \tilde{x}| \leq \varepsilon$, and $y \in \mathbb{S}^{N-1}, \tilde{y} \in \mathcal{M}$ so that $|y - \tilde{y}| \leq \varepsilon$. We have

$$\begin{aligned} |\langle Ax, y\rangle - \langle A\tilde{x}, \tilde{y}\rangle| &= |\langle Ax, y\rangle - \langle Ax, \tilde{y}\rangle + \langle Ax, \tilde{y}\rangle - \langle A\tilde{x}, \tilde{y}\rangle| \\ &\leq |\langle Ax, y - \tilde{y}\rangle| + |\langle A(x - \tilde{x}), \tilde{y}\rangle| \\ &\leq 2\|A\| \cdot \varepsilon \\ &= 2\varepsilon \cdot \sup_{x \in \mathbb{S}^{n-1}, y \in \mathbb{S}^{N-1}} \langle Ax, y\rangle, \end{aligned} \tag{20}$$

which implies (19).

Going back to our aim, we have

$$\begin{aligned}
\mathbb{P}\left(\|A\| \geq C\cdot(\sqrt{N}+t)\right) &= \mathbb{P}\left(\sup_{x\in\mathbb{S}^{n-1},y\in\mathbb{S}^{N-1}}\langle Ax,y\rangle \geq C\cdot(\sqrt{N}+t)\right)\\
&\leq \mathbb{P}\left(\sup_{x\in\mathcal{N},y\in\mathcal{M}}\langle Ax,y\rangle \geq \frac{C}{2}\cdot(\sqrt{N}+t)\right)\\
&= \mathbb{P}\left(\bigcup_{x\in\mathcal{N},y\in\mathcal{M}}\{\langle Ax,y\rangle \geq \tilde{C}\cdot(\sqrt{N}+t)\}\right)\\
&\leq \sum_{x\in\mathcal{N},y\in\mathcal{M}}\mathbb{P}\left(\langle Ax,y\rangle \geq \tilde{C}\cdot(\sqrt{N}+t)\right)\\
&\leq \sup_{x\in\mathcal{N},y\in\mathcal{M}}\mathbb{P}\left(\langle Ax,y\rangle \geq \tilde{C}\cdot(\sqrt{N}+t)\right)\cdot 9^{n+N}.
\end{aligned}$$

The first inequality holds because of the inequality (19) and plugging $1-2\varepsilon=\frac{1}{2}$.

It remains to bound the term $\mathbb{P}\left(\langle Ax,y\rangle \geq \tilde{C}\cdot(\sqrt{N}+t)\right)$ for fixed $x\in\mathbb{S}^{n-1}$, $y\in\mathbb{S}^{N-1}$. Note that

$$\mathbb{P}\left(\langle Ax,y\rangle \geq \tilde{C}\cdot(\sqrt{N}+t)\right) = \mathbb{P}\left(\sum_{ij}a_{ij}\cdot x_jy_i \geq \tilde{C}\cdot(\sqrt{N}+t)\right). \tag{21}$$

We claim that $\sum_{i=1}^{N}\sum_{j=1}^{n}a_{ij}x_jy_i$ is sub-Gaussian. Indeed, the bound for a sum of independent mean-zero sub-Gaussian random variables gives

$$\begin{aligned}
\left\|\sum_{i=1}^{N}\sum_{j=1}^{n}a_{ij}x_jy_i\right\|_{\psi_2}^2 &\leq C'\sum_{i=1}^{N}\sum_{j=1}^{n}|x_jy_i|^2\|a_{ij}\|_{\psi_2}^2\\
&\leq C'\left(\sum_{j=1}^{n}x_j^2\right)\left(\sum_{i=1}^{N}y_i^2\right) = C'.
\end{aligned} \tag{22}$$

Here we used $K=1$ and the fact that x,y are unit vectors. As a result, we have

$$\mathbb{P}\left(\langle Ax,y\rangle \geq \tilde{C}\cdot(\sqrt{N}+t)\right) \leq 2e^{-C''\cdot(\sqrt{N}+t)^2}. \tag{23}$$

Combining the previous results we have

$$\mathbb{P}\left(\|A\| \geq C(\sqrt{N}+t)\right) \leq 9^{n+N}\cdot 2e^{-C''\cdot(\sqrt{N}+t)^2} \leq 2e^{-t^2}, \tag{24}$$

provided that $C>0$ is selected sufficiently large, using $n\leq N$. Rescaling restores the factor K in the conclusion. □

Corollary 4.5. *Under the assumptions of Theorem 4.4, we have* $\|A\| \le CK\sqrt{N}$ *with probability at least* $1 - 2e^{-N}$*, by taking* $t = \sqrt{N}$*. Integrating the tail bound in that theorem also gives* $\mathbb{E}\|A\| \le \widetilde{C}K\sqrt{N}$*.*

Corollary 4.6. *Let* A *be an* $n \times n$ *symmetric random matrix whose entries on and above the diagonal are independent, mean zero, and have sub-Gaussian norms at most* K*. Then for all* $t \ge 0$*, we have*

$$\|A\| \le CK(\sqrt{n} + t)$$

with probability at least $1 - 4e^{-t^2}$*.*

Proof. Homework! □

4.2 Two-sided singular-value bounds for tall matrices

Recall the following notion:

Definition 4.7. $X \in \mathbb{R}^n$ is called an isotropic random vector in $\mathbb{R}^n$ if $\mathbb{E}X = 0$, and for all $\theta \in \mathbb{S}^{n-1}$,

$$\mathbb{E}\langle X, \theta\rangle^2 = 1.$$

Note some similarity of this notion and the notion of sub-Gaussian random vectors – the vectors for which $\langle X, \theta\rangle$ is sub-Gaussian for every $\theta \in \mathbb{S}^{n-1}$.

Theorem 4.8 (Two-sided bound on sub-Gaussian matrices)**.** *Let* A *be an* $N \times n$ *matrix whose rows* A_i *are independent, mean zero, sub-gaussian isotropic random vectors in* $\mathbb{R}^n$*. Then for any* $t \ge 0$ *we have*

$$\sqrt{N} - CK^2(\sqrt{n} + t) \le \sigma_n(A) \le \sigma_1(A) \le \sqrt{N} + CK^2(\sqrt{n} + t) \tag{25}$$

with probability at least $1 - 2e^{-t^2}$*. Here* $K = \max_i \|A_i\|_{\psi_2}$*.*

Theorem 4.8 complements Theorem 4.4. Under the additional isotropy assumption, it gives the following improvements:

- it allows dependence within each row. Examples include independent rows sampled uniformly from $\sqrt{n}\mathbb{S}^{n-1}$ and orthogonal projections of isotropic sub-Gaussian vectors onto a fixed subspace, expressed in an orthonormal basis.
- it is a two-sided bound for all singular values (rather than just an upper bound for only $\sigma_1(A)$.)
- If $N \gg n$, the bound is more precise, recovering the constant 1 in front of $\sqrt{N}$. Note also that the taller the matrix, the lesser is the impact of the sub-Gaussian constant.

However, note that Theorem 4.8 gives a positive lower bound at a fixed high-probability level only when $N \ge C(K)n$ *with* $C(K)$ *sufficiently large.* This is very important to note.

We start the proof by pointing out

Claim 4.9. *Let A be an $N \times n$ matrix and $\delta > 0$. Suppose that*

$$\|A^\top A - I_n\| \le \max(\delta, \delta^2). \tag{26}$$

Then

$$(1-\delta)|x| \le |Ax| \le (1+\delta)|x| \quad \text{for all } x \in \mathbb{R}^n. \tag{27}$$

The proof of this elementary fact is left as a homework.

Therefore, to establish Theorem 4.8, it is enough to prove that

$$\|\frac{1}{N} A^T A - I_n\| \le K^2 \max(\delta, \delta^2)$$

with high probability, where $\delta = C\left(\sqrt{\frac{n}{N}} + \frac{t}{\sqrt{N}}\right)$. Indeed, isotropy and the definition of the sub-Gaussian norm imply $K \ge 1$, and

$$K^2 \max(\delta, \delta^2) \le \max(K^2\delta, (K^2\delta)^2).$$

Apply Claim 4.9 to $A/\sqrt{N}$ with parameter $K^2\delta$. The proof will be done via the epsilon-net argument. As a first step, we shall show the point-wise bound, which relies on the Bernstein inequality which we proved before.

Lemma 4.10. *Let A be as in Theorem 4.8. Fix $s > 0$ and $x \in \mathbb{S}^{n-1}$. Then*

$$\mathbb{P}\left\{\left|\frac{1}{N}|Ax|^2 - 1\right| \ge s\right\} \le 2e^{-CN\min\{\frac{s^2}{K^4}, \frac{s}{K^2}\}}. \tag{28}$$

Proof. Consider the random vector

$$Ax = \begin{pmatrix} \langle A^T e_1, x\rangle \\ \langle A^T e_2, x\rangle \\ \vdots \\ \langle A^T e_N, x\rangle \end{pmatrix}. \tag{29}$$

Note that the coordinates of Ax are independent by assumption. Also, by assumption, since $x \in \mathbb{S}^{n-1}$, we have $\|\langle A^T e_i, x\rangle\|_{\psi_2} \le K$. Now, isotropicity implies that $\mathbb{E}\langle A^T e_i, x\rangle^2 = 1$. Thus, $\frac{1}{N}|Ax|^2 - 1$ is a sum of mean-zero sub-exponential independent random variables with sub-exponential norms bounded by $\frac{CK^2}{N}$. An application of Bernstein's inequality (Theorem 3.33) thus finishes the proof. □

Proof of Theorem 4.8. Recall that there exists a 1/4-net $\mathcal{N} \subset \mathbb{S}^{n-1}$ with $\#\mathcal{N} \le 9^n$ such that for any $x \in \mathbb{S}^{n-1}$ there exists a $y \in \mathcal{N}$ such that $|x - y| \le \frac{1}{4}$, and therefore,

$$\left\|\frac{1}{N} A^\top A - I_n\right\| \le 2 \max_{x \in \mathcal{N}} \left|\frac{1}{N}|Ax|^2 - 1\right|. \tag{30}$$

Here the matrix is symmetric, so its norm is the supremum of the absolute value of its quadratic form on the unit sphere. Approximating both occurrences of x by the same net point gives the factor 2, as in the proof of Theorem 4.4. Therefore, using the union bound as before, we see:

$$\begin{aligned}\mathbb{P}\left\{\left\|\frac{1}{N}A^TA - I_n\right\| \geq K^2\max(\delta,\delta^2)\right\} &\leq \mathbb{P}\left\{\max_{x\in\mathcal{N}}\left|\frac{1}{N}|Ax|^2 - 1\right| \geq \frac{K^2}{2}\max(\delta,\delta^2)\right\} \\ &\leq 2\cdot 9^n e^{-cN\delta^2} \leq 2e^{-t^2},\end{aligned} \tag{31}$$

where $\delta = C(\sqrt{n/N} + t/\sqrt{N})$ and C is sufficiently large. For the second inequality, apply Lemma 4.10 with $s = \frac{K^2}{2}\max(\delta,\delta^2)$; then $\min(s^2/K^4, s/K^2) \geq c\delta^2$. The last inequality absorbs the factor 9^n. Claim 4.9, applied as explained above, finishes the proof. □

4.3 Matrix Bernstein Inequality

Theorem 4.11 (Matrix Bernstein Inequality)**.** *Let $X_1, X_2, \cdots, X_N$ be independent mean-zero $n\times n$ symmetric random matrices. Suppose $\|X_i\| \leq K$ for all $i = 1,2,\ldots,N$ almost surely (i.e. with probability 1). Then, for all $t \geq 0$,*

$$\mathbb{P}\left\{\left\|\sum_{i=1}^N X_i\right\| \geq t\right\} \leq 2n\cdot e^{-\frac{t^2/2}{\sigma^2 + Kt/3}}, \tag{32}$$

where $\sigma^2 = \|\sum_{i=1}^N \mathbb{E}X_i^2\|$. In particular,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^N X_i\right\| \geq t\right\} \leq 2n\cdot e^{-C\min\{\frac{t^2}{\sigma^2},\frac{t}{K}\}}. \tag{33}$$

Note that $\|\sum_{i=1}^N X_i\|$ is the norm of a random matrix and $\|\sum_{i=1}^N \mathbb{E}X_i^2\|$ is a norm of a deterministic matrix.

Is it possible to deduce Corollary 4.6 directly from Theorem 4.11? Let us examine the natural decomposition.

Suppose that

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

is symmetric, and decompose A into $n(n+1)/2$ independent symmetric matrices, one for

each diagonal entry and each pair of symmetric off-diagonal entries:

$$X_1 = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \; X_2 = \begin{pmatrix} 0 & a_{12} & 0 & \cdots & 0 \\ a_{21} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$X_3 = \begin{pmatrix} 0 & 0 & a_{13} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ a_{31} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \; \ldots$$

The summands are independent and mean zero, but sub-Gaussian entries need not be almost surely bounded, so the displayed version of matrix Bernstein does not apply directly. Even if $|a_{ij}| \le K$ almost surely, it gives $\sigma^2 \le nK^2$ and hence

$$\mathbb{P}\left\{ \|A\| \ge CK\big(\sqrt{n(\log(2n)+t)} + \log(2n) + t\big) \right\} \le e^{-t}, \qquad t \ge 0.$$

Thus this decomposition introduces a logarithmic loss and does not recover Corollary 4.6. Checking the variance bound and this consequence is a useful exercise.

Theorem 4.12 (Lieb's Inequality)**.** *Let H be an $n \times n$ symmetric matrix. Consider $f(X) = \operatorname{tr}(e^{H+\log X})$ to be a function of the symmetric positive definite matrix X. Then f is concave on this space:*

$$\operatorname{tr}(e^{H+\log \frac{X+Y}{2}}) \ge \frac{1}{2}\left(\operatorname{tr}(e^{H+\log X}) + \operatorname{tr}(e^{H+\log Y})\right) \tag{34}$$

Recall Jensen's Inequality (which works also for matrices): if f is concave, then $\mathbb{E}f(X) \le f(\mathbb{E}X)$. As a result, we get

Corollary 4.13 (Lieb's Inequality + Jensen's Inequality)**.** *Let H be a fixed $n \times n$ symmetric matrix and Z be a random $n \times n$ symmetric matrix. Assume the relevant expectations are finite. Then we have*

$$\mathbb{E}\left(\operatorname{tr}\left(e^{H+Z}\right)\right) \le \operatorname{tr}\left(e^{H+\log \mathbb{E}e^{Z}}\right) \tag{35}$$

We are now ready to prove Theorem 4.11.

Proof of Theorem 4.11. Denote $S = \sum_{i=1}^{N} X_i$. Since S is symmetric,

$$\|S\| = \max_{1 \le i \le n} |\lambda_i(S)| = \max\{\lambda_{\max}(S), \lambda_{\max}(-S)\}. \tag{36}$$

We first estimate the upper tail of $\lambda_{\max}(S)$, and then apply the same argument to $-S$. For $\lambda > 0$, the Chernoff argument gives

$$\begin{aligned}\mathbb{P}\{\lambda_{\max}(S) \geq t\} &\leq e^{-\lambda t}\,\mathbb{E}\,e^{\lambda\lambda_{\max}(S)} \\ &= e^{-\lambda t}\,\mathbb{E}\,\lambda_{\max}(e^{\lambda S}) \leq e^{-\lambda t}\,\mathbb{E}\,\operatorname{tr}(e^{\lambda S}).\end{aligned}$$

The equality uses the definition of functions of a symmetric matrix, and the last inequality uses that $e^{\lambda S}$ is positive definite, so its largest eigenvalue is bounded by its trace.

Applying Lieb's inequality conditionally, and using independence at each step, yields

$$\begin{aligned}\mathbb{E}\,\operatorname{tr}(e^{\lambda S}) &\leq \mathbb{E}\,\operatorname{tr}\left(e^{\sum_{i=1}^{N-1}\lambda X_i + log\,\mathbb{E}\,e^{\lambda X_N}}\right) \\ &\leq \operatorname{tr}\left(e^{\sum_{i=1}^{N}\log\mathbb{E}\,e^{\lambda X_i}}\right).\end{aligned}$$

For example, in the first step we condition on $X_1, \ldots, X_{N-1}$ and apply the preceding corollary with $H = \sum_{i=1}^{N-1}\lambda X_i$ and $Z = \lambda X_N$. We must keep the sum of matrix logarithms inside the exponential: the matrices $\mathbb{E}\,e^{\lambda X_i}$ need not commute.

It remains to bound the individual matrix moment generating functions.

Lemma 4.14 (Homework)**.** *Let X be a symmetric random matrix with $\mathbb{E}\,X = 0$ and $\|X\| \leq K$ almost surely. If $|\lambda| < 3/K$, then, in the positive-semidefinite order,*

$$\mathbb{E}\,e^{\lambda X} \leq e^{g(\lambda)\,\mathbb{E}\,X^2}, \qquad g(\lambda) = \frac{\lambda^2/2}{1 - |\lambda|K/3}.$$

Remark 4.15. *Here $U \leq V$ means that $V - U$ is positive semidefinite, that is, $\langle (V - U)x, x\rangle \geq 0$ for every x.*

The matrix logarithm is monotone in this order, and the trace of the matrix exponential is also monotone. Therefore, for $0 < \lambda < 3/K$,

$$\begin{aligned}\mathbb{E}\,\operatorname{tr}(e^{\lambda S}) &\leq \operatorname{tr}\exp\left(g(\lambda)\sum_{i=1}^{N}\mathbb{E}\,X_i^2\right) \\ &\leq n\exp\left(g(\lambda)\sigma^2\right).\end{aligned}$$

Consequently,

$$\mathbb{P}\{\lambda_{\max}(S) \geq t\} \leq n\exp\left(-\lambda t + \frac{\lambda^2\sigma^2/2}{1 - \lambda K/3}\right).$$

For $\sigma^2 > 0$ and $t > 0$, take $\lambda = t/(\sigma^2 + Kt/3)$. This gives

$$\mathbb{P}\{\lambda_{\max}(S) \geq t\} \leq n\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right).$$

The same estimate holds for $-S$, since the matrices $-X_i$ satisfy the same assumptions. A union bound and (36) give the factor $2n$ in the theorem. If $\sigma^2 = 0$, every $X_i = 0$ almost surely, and the conclusion for $t > 0$ is immediate; the bound at $t = 0$ is trivial. □

4.4 Non-asymptotic bounds for the smallest singular value of random matrices

4.4.1 General discussion about the smallest singular value of a random matrix

Theorem 4.8 implies that if an $N \times n$ matrix has independent, mean-zero, variance-one entries with sub-Gaussian norms bounded by K, then, with probability at least $1 - 2e^{-t^2}$,

$$\sqrt{N} - CK^2(\sqrt{n} + t) \le \sigma_n(A) \le \cdots \le \sigma_1(A) \le \sqrt{N} + CK^2(\sqrt{n} + t).$$

The lower bound is positive when N is sufficiently large compared with $K^4(n + t^2)$. Upper bounds for the operator norm require control of the upper tails of the entries: sub-Gaussian assumptions give exponential probability bounds, while fourth-moment assumptions naturally arise in estimates at the scale $\sqrt{N}$ for i.i.d. matrices; see, for example, Litvak and Spektor [25]. In sharp contrast, much weaker assumptions can suffice for lower bounds on

$$\sigma_n(A) = \inf_{x \in \mathbb{S}^{n-1}} |Ax|.$$

This phenomenon was developed in works of Tikhomirov [51; 52], Rebrova and Tikhomirov [33], Livshyts [26], Guédon, Litvak and Tatarko [13], Livshyts, Tikhomirov and Vershynin [27], and others.

Let us investigate how the bound of the type

$$\sigma_n(A) \ge \clubsuit, \tag{37}$$

can be proven (on average or with high probability), and what might be required of the random matrix A for this. For a square $n \times n$ matrix A, the condition $\sigma_n(A) = 0$ is equivalent to A being non-invertible. Therefore, in the case $N = n$, the condition (37) means that the matrix A is "well invertible", and having such information about a random matrix could be valuable for applications in various situations.

Let us start by making a naive attempt to use the net argument to bound $\sigma_n(A)$ from below. For $0 < \varepsilon < 1$, let $\mathcal{N} \subset \mathbb{S}^{n-1}$ be an ε-net of size at most $\left(\frac{3}{\varepsilon}\right)^n$. Then for all $x \in \mathbb{S}^{n-1}$, there exists $y \in \mathcal{N}$ such that $|x - y| \le \varepsilon$. That implies that $|A(x-y)| \le \|A\| \cdot \varepsilon$. By triangle inequality, $|Ax| \ge |Ay| - \|A\| \cdot \varepsilon$. Taking infimum on both sides, we get

$$\inf_{x \in \mathbb{S}^{n-1}} |Ax| \ge \inf_{y \in \mathcal{N}} |Ay| - \|A\| \cdot \varepsilon.$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left(\sigma_n(A) \le \clubsuit\right) &= \mathbb{P}\left(\inf_{x \in \mathbb{S}^{n-1}} |Ax| \le \clubsuit\right) \le \mathbb{P}\left(\inf_{y \in \mathcal{N}} |Ay| \le \clubsuit + \|A\| \cdot \varepsilon\right) \\
&= \mathbb{P}\left(\bigcup_{y \in \mathcal{N}} \left\{|Ay| \le \clubsuit + \|A\| \cdot \varepsilon\right\}\right) \\
&\le \#\mathcal{N} \cdot \sup_{y \in \mathcal{N}} \mathbb{P}\left(|Ay| \le \clubsuit + \|A\| \cdot \varepsilon\right).
\end{aligned}$$

In case the assumptions on the random matrix allow us to have a good control of $\|A\|$ then this is a promising start! However, as we mentioned earlier, one should not need to make such strong assumptions in principle. Thus we are going to apply Theorem 2.23 (about a more involved net argument) instead, in order to aim for a more general result. Theorem 2.23 was valid for all deterministic matrices, so we point out the following Corollary for random matrices:

Corollary 4.16. *Suppose A is any random matrix, with $\mathbb{E}\,\|A\|_{HS}^2 < \infty$. Then for any $0 < \varepsilon < 1/2$, there is a deterministic net $\mathcal{N} \subset \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n$ with $\#\mathcal{N} \leq (C/\varepsilon)^n$ such that with probability at least 0.9 we have for all $x \in \mathbb{S}^{n-1}$ there is some $y \in \mathcal{N}$ such that*

$$|A(x-y)| \leq \sqrt{10}\varepsilon \cdot \frac{\sqrt{\mathbb{E}\,\|A\|_{HS}^2}}{\sqrt{n}}.$$

Proof. Take the net $\mathcal{N}$ from Theorem 2.23. Then for all $x \in \mathbb{S}^{n-1}$ there is some $y \in \mathcal{N}$ such that

$$|A(x-y)| \leq \varepsilon \cdot \frac{\|A\|_{HS}}{\sqrt{n}}.$$

If $\mathbb{E}\,\|A\|_{HS}^2 = 0$, then $A = 0$ almost surely and the assertion is immediate. Otherwise, Markov's inequality gives

$$\mathbb{P}\Big(\|A\|_{HS}^2 \leq 10\,\mathbb{E}\,\|A\|_{HS}^2\Big) = 1 - \mathbb{P}\Big(\|A\|_{HS}^2 > 10\,\mathbb{E}\,\|A\|_{HS}^2\Big) \geq 1 - 0.1 = 0.9.$$

This implies that with probability at least 0.9, for all $x \in \mathbb{S}^{n-1}$ there is some $y \in \mathcal{N}$ such that

$$|A(x-y)| \leq \varepsilon \cdot \frac{\sqrt{10\,\mathbb{E}\,\|A\|_{HS}^2}}{\sqrt{n}}.$$

□

4.4.2 Small-ball assumptions and tensorization

In order to apply any kind of net argument, we need to be able to upper bound $\mathbb{P}(|Ay| \leq \clubsuit)$ for a fixed $y \in \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n$. To this end, we shall now have a discussion about a small ball (or anti-concentration) assumption for a random variable and for a random vector.

Definition 4.17. Suppose ξ is a random variable. We say that it satisfies a small ball (or an anti-concentration) assumption if for any $z \in \mathbb{R}$, $\mathbb{P}(|\xi - z| \leq a) \leq b$ for some $a > 0$, $b \in (0,1)$.

In other words, ξ does not concentrate around some point z ("small ball") with too high probability. Here are some examples:

- if ξ has a bounded density f (say, bounded by some constant K), then for any interval I with length $2a$,
$$\mathbb{P}(\xi \in I) = \int_I f \leq K|I| = 2aK.$$
One may let $a = \frac{1}{4K}$, $b = \frac{1}{2}$, and then for any $z \in \mathbb{R}$, $\mathbb{P}(|\xi - z| \leq a) \leq b$.

- Consider the symmetric Bernoulli
$$\xi = \begin{cases} 1, \text{ with probability } \frac{1}{2}, \\ -1, \text{ with probability } \frac{1}{2}. \end{cases}$$
Then for any $z \in \mathbb{R}$, $\mathbb{P}(|\xi - z| \leq 0.9) \leq \frac{1}{2}$.

- If ξ is mean zero, has variance one, and has sub-Gaussian norm at most K, then it satisfies a uniform small-ball estimate with a, b depending only on K (an exercise using symmetrization and the Paley–Zygmund inequality). The nondegeneracy assumption is essential: a constant random variable is bounded and sub-Gaussian, but is not anti-concentrated.

We shall rely heavily on the following Lemma which appears in Rudelson, Vershynin [36].

Lemma 4.18 (Tensorization Lemma)**.** *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent non-negative random variables. Fix $\varepsilon_0 > 0$, $K > 0$, and suppose that for any $\varepsilon \geq \varepsilon_0$,*
$$\mathbb{P}(\xi_k \leq \varepsilon) \leq K\varepsilon, k = 1, 2, \ldots, n.$$
Then, for every $\varepsilon \geq \varepsilon_0$ and an absolute constant $C > 0$,
$$\mathbb{P}\left(\sum_{j=1}^n \xi_j^2 \leq \varepsilon^2 n\right) \leq (CK\varepsilon)^n.$$

In other words, independent one-dimensional small-ball bounds yield a bound for the Euclidean length of the vector.

Proof. Assume $\varepsilon \geq \varepsilon_0$. Using Markov's inequality and the independence of ξ_j's we have:
$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^n \xi_j^2 \leq \varepsilon^2 n\right) &= \mathbb{P}\left(\exp\left\{-\frac{1}{\varepsilon^2}\sum_{j=1}^n \xi_j^2\right\} \geq e^{-n}\right) \\ &\leq e^n \cdot \mathbb{E}\exp\left\{-\frac{1}{\varepsilon^2}\sum_{j=1}^n \xi_j^2\right\} \\ &= e^n \cdot \prod_{j=1}^n \mathbb{E}\exp\left\{-\frac{1}{\varepsilon^2}\xi_j^2\right\}. \end{aligned}$$

Writing out the expectation and changing variables $t = e^{-s^2}$, we get

$$\begin{aligned}\mathbb{E}\exp\left\{-\frac{1}{\varepsilon^2}\xi_j^2\right\} &= \int_0^1 \mathbb{P}\left(\exp\left\{-\frac{1}{\varepsilon^2}\xi_j^2\right\} \geq t\right) dt \\ &= \int_0^\infty 2se^{-s^2}\mathbb{P}\left(\xi_j \leq \varepsilon s\right) ds.\end{aligned}$$

Since $\varepsilon \geq \varepsilon_0$, if $s \geq 1$, then $\varepsilon s \geq \varepsilon_0$ and $\mathbb{P}\left(\xi_j \leq \varepsilon s\right) \leq Ks\varepsilon$. If $s \leq 1$, then

$$\mathbb{P}\left(\xi_j \leq \varepsilon s\right) \leq \mathbb{P}\left(\xi_j \leq \varepsilon\right) \leq K\varepsilon.$$

So we can break the integral into two parts and get, for some constant $C_1 > 0$:

$$\int_0^\infty 2se^{-s^2}\mathbb{P}\left(\xi_j \leq \varepsilon s\right) ds \leq \int_0^1 2K\varepsilon s e^{-s^2} ds + \int_1^\infty 2K\varepsilon s^2 e^{-s^2} ds \leq C_1 K\varepsilon.$$

Combining all of the above and letting $C = e \cdot C_1$, we get

$$\mathbb{P}\left(\sum_{j=1}^n \xi_j^2 \leq \varepsilon^2 n\right) \leq e^n \cdot (C_1 K\varepsilon)^n = (CK\varepsilon)^n.$$

□

Definition 4.19 (High-dimensional version of the small-ball assumption)**.** A random vector $X \in \mathbb{R}^n$ is uniformly anti-concentrated if there exist $a > 0$ and $b \in (0,1)$ such that

$$\sup_{z\in\mathbb{R}} \mathbb{P}\big(|\langle X, \theta\rangle - z| \leq a\big) \leq b \qquad \text{for every } \theta \in \mathbb{S}^{n-1}.$$

For the next matrix lemma, the weaker bound with $z = 0$ already suffices.

Claim 4.20 (Homework)**.** *Let $X \in \mathbb{R}^n$ have independent coordinates satisfying*

$$\sup_{z\in\mathbb{R}} \mathbb{P}(|X_i - z| \leq a) \leq b < 1, \qquad i = 1, \ldots, n.$$

Then there exist $a' > 0$ and $b' \in (0,1)$, depending only on a, b, such that

$$\sup_{z\in\mathbb{R}} \mathbb{P}\big(|\langle X, \theta\rangle - z| \leq a'\big) \leq b' \qquad \textit{for every } \theta \in \mathbb{S}^{n-1}.$$

This is a standard consequence of Rogozin's theorem; see [33] and the discussion below.

The following lemma is a fixed-scale version of the tensorization principle in Lemma 4.18, whose proof we leave as a homework.

Lemma 4.21. *Suppose A is an $N \times n$ random matrix with independent rows $A^\top e_i$, $i = 1, 2, \cdots, N$, and suppose for any $\theta \in \mathbb{S}^{n-1}$,*

$$\mathbb{P}\Big(\big|\langle A^\top e_i, \theta\rangle\big| < a\Big) < b.$$

Then for any $x \in \mathbb{S}^{n-1}$,

$$\mathbb{P}\Big(|Ax| \le c\sqrt{N}\Big) \le e^{-c_1 N},$$

where $c, c_1 > 0$ only depend on a and b.

We point out also the following

Corollary 4.22 (Claim 4.20 combined with Lemma 4.21)**.** *Suppose A is an $N \times n$ random matrix with independent entries a_{ij}. If for all i, j and some $a > 0$, $b \in (0, 1)$,*

$$\sup_{z\in\mathbb{R}} \mathbb{P}(|a_{ij} - z| \le a) \le b,$$

then for any $x \in \mathbb{S}^{n-1}$,

$$\mathbb{P}\Big(|Ax| \le c\sqrt{N}\Big) \le e^{-c_1 N},$$

where $c, c_1 > 0$ only depend on a and b.

4.4.3 The smallest singular value of tall random matrices

We are finally ready to establish our first bound on the smallest singular value of a random matrix. So far, we will require the strong assumption that $N \ge Cn$. The statement below appears as Proposition 1 in Livshyts [26].

Proposition 4.23 (Tall matrices with possibly dependent columns)**.** *Let A be an $N \times n$ random matrix whose rows $A^T e_i$ are independent. Given $a > 0$ and $b \in (0, 1)$, suppose that for every row $A^\top e_i$ and any $\theta \in \mathbb{S}^{n-1}$,*

$$\mathbb{P}\Big(|\langle A^\top e_i, \theta\rangle| < a\Big) < b.$$

Also assume that

$$\mathbb{E}\,\|A\|_{HS}^2 = \sum_{i,j} \mathbb{E}\, a_{ij}^2 \le KNn$$

for some $K > 0$. Then there exist $C_1 \ge 1$, and $c_2, c_3 > 0$ which only depend on a, b and K such that whenever $N \ge C_1 n$ then

$$\mathbb{P}\Big(\sigma_n(A) \le c_2\sqrt{N}\Big) \le 0.2.$$

Consequently, $\mathbb{E}\,\sigma_n(A) \ge c_3\sqrt{N}$.

Proof. By Lemma 4.21, there are $c_0, c_1 > 0$, depending only on a, b, such that

$$\sup_{|u|=1} \mathbb{P}\{|Au| \le c_0\sqrt{N}\} \le e^{-c_1 N}.$$

Choose $0 < \varepsilon < 1/2$ so small that $\sqrt{10K}\varepsilon \le c_0/4$, and apply Corollary 4.16. It gives a deterministic net $\mathcal{N}$ in the stated annulus, of cardinality at most $(C/\varepsilon)^n$, such that, on an event of probability at least 0.9, every $x \in \mathbb{S}^{n-1}$ has a net point y with

$$|A(x-y)| \le \sqrt{10K}\varepsilon\sqrt{N} \le \frac{c_0}{4}\sqrt{N}.$$

Take $c_2 = c_0/4$. If $\sigma_n(A) \le c_2\sqrt{N}$ on this event, some $y \in \mathcal{N}$ satisfies $|Ay| \le (c_0/2)\sqrt{N}$. Since $|y| \ge 1/2$, the pointwise bound applies to $y/|y|$ and gives

$$\begin{aligned}\mathbb{P}\{\sigma_n(A) \le c_2\sqrt{N}\} &\le 0.1 + \sum_{y\in\mathcal{N}} \mathbb{P}\{|Ay| \le (c_0/2)\sqrt{N}\} \\ &\le 0.1 + (C/\varepsilon)^n e^{-c_1 N} \le 0.2,\end{aligned}$$

provided $N \ge C_1 n$ with $C_1 = C_1(a, b, K)$ sufficiently large. Finally, $\mathbb{E}\,\sigma_n(A) \ge 0.8c_2\sqrt{N}$. □

Proposition 4.23 should be compared with Theorem 4.8: while it only discusses the smallest singular value, the assumptions on the matrix are a lot less demanding. Keep in mind also that Theorem 4.8 implies sharpness of the estimate in the Proposition 4.23 at least in many important situations. In both cases the assumption $N \ge Cn$ is crucial. Soon we will see that the situation is quite different for square matrices, and the related results are far more complicated.

Let us formulate a corollary of Proposition 4.23 combined with Claim 4.20:

Corollary 4.24. *Let A be an $N \times n$ random matrix with the following properties:*

1. *All entries are independent;*
2. *The entries are Uniformly Anti-Concentrated (UAC), that is, there exist $a > 0$ and $b \in (0,1)$ such that $\sup_{z\in\mathbb{R}} \mathbb{P}(|a_{ij} - z| \le a) \le b$;*
3. *$\mathbb{E}\,\|A\|_{HS}^2 \le K \cdot Nn$.*

If $N \ge Cn$ for an appropriate $C \ge 1$ that depends on a, b, K, then $\mathbb{E}\,\sigma_n(A) \ge c_1\sqrt{N}$, where $c_1 > 0$ also depends only on a, b, K.

What if we want this estimate to hold with high probability rather than on average, that is, would it be possible to prove

$$\mathbb{P}(\sigma_n(A) \le C\sqrt{N}) \le e^{-cn}$$

in place of

$$\mathbb{P}(\sigma_n(A) \le C\sqrt{N}) \le 0.1?$$

Turns out, this is possible! Analyzing the proof, one may note that the 0.1 error in the probability estimate came from the fact that our net only works with constant probability (as was obtained in Corollary 4.16 using Markov's inequality). But it is possible to construct a net with similar properties which would in fact work well with high probability! We discuss this construction below.

4.4.4 A net construction which works with high probability for matrices with independent columns

Theorem 4.25 (Livshyts [26]). *Fix $n \in \mathbb{N}$ and $\varepsilon \in (0, 0.1)$. There is a deterministic net $\mathcal{N} \subset \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n$, of cardinality at most $(C/\varepsilon)^n$, such that for any random $N \times n$ matrix A with independent columns and finite $\mathbb{E}\,\|A\|_{HS}^2$, the following holds with probability at least $1 - e^{-5n}$: for every $x \in \mathbb{S}^{n-1}$, there is $y \in \mathcal{N}$ satisfying*

$$|A(x-y)| \le \frac{C\varepsilon}{\sqrt{n}}\sqrt{\mathbb{E}\,\|A\|_{HS}^2}.$$

Here $C > 0$ is an absolute constant.

One should compare Theorem 4.25 to Corollary 4.16. On one hand, the probability in the new result is much larger: $1 - e^{-5n} >> 0.9$ when $n \to \infty$. On the other hand, we pay by requiring the matrix to have independent columns.

Before we discuss the proof of Theorem 4.25, we state the following improvement of Corollary 4.24:

Corollary 4.26. *Let A be an $N \times n$ random matrix whose entries are independent and satisfy*

$$\sup_{z \in \mathbb{R}} \mathbb{P}(|a_{ij} - z| \le a) \le b < 1, \qquad \mathbb{E}\,\|A\|_{HS}^2 \le KNn.$$

There are $C \ge 1$ and $c_1, c_2 > 0$, depending only on a, b, K, such that if $N \ge Cn$, then

$$\mathbb{P}\{\sigma_n(A) \le c_1\sqrt{N}\} \le e^{-c_2 n}.$$

Corollary 4.26 follows from Theorem 4.25 verbatim as Corollary 4.24 follows from Corollary 4.16, so we leave the details to the reader.

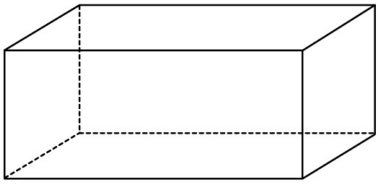
4.5 Proof of Theorem 4.25.

Recall that earlier, we constructed a net on the sphere which was a subset of the cubic lattice net. Here, inspired by Rebrova and Tikhomirov [33], we will instead cover the sphere by coordinate boxes, that is, sets of the form $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$.

Step 1: Cover the sphere by coordinate boxes and consider the associated random rounding. Up to translation, a coordinate box P is determined by the lengths of its sides $\alpha_1, \alpha_2, \ldots, \alpha_n \in [0,1]$, or in other words, a coordinate box

$$P = [0, \alpha_1] \times ... \times [0, \alpha_n]$$

is determined by the vector $\alpha = (\alpha_1, \ldots \alpha_n)$.



Fix also a parameter $\kappa \geq 1$.
Consider

$$\Omega_\kappa = \{\alpha \in \mathbb{R}^n : \alpha_i \in [0,1] \ \forall i \text{ and } \prod_{i=1}^{n} \alpha_i \geq \kappa^{-n}\}$$

This is the "set of admissible parallelepipeds", a set of coordinate boxes that fit inside the cube with volume at least κ^{-n}.

Lemma 4.27. *Fix $\kappa \geq 1, \varepsilon \in (0, 0.1), \alpha \in \Omega_\kappa$, and consider $P_\alpha = [0, \alpha_1] \times \cdots \times [0, \alpha_n]$. Then, there exist $y_1, \ldots, y_m \in 1.4B_2^n$ so that $\mathbb{S}^{n-1} \subset \bigcup_{i=1}^m (y_i + \frac{\varepsilon}{\sqrt{n}} P_\alpha)$ and $m \leq (\frac{10}{\varepsilon}\kappa)^n$.*

This fact is proved by considering

$$\mathcal{F}_\alpha = \left(\frac{\varepsilon\alpha_1}{\sqrt{n}}\mathbb{Z} \times ... \times \frac{\varepsilon\alpha_n}{\sqrt{n}}\mathbb{Z}\right) \cap \left(\frac{3}{2}B_2^n \setminus \frac{1}{2}B_2^n\right), \tag{38}$$

and taking the lower-left vertices y_i of the lattice boxes that meet the sphere. Every such box has diameter at most ε, so all its vertices lie in the displayed annulus. The boxes have disjoint interiors and volume $\varepsilon^n \prod_i \alpha_i / n^{n/2} \geq (\varepsilon/(\kappa\sqrt{n}))^n$; a volume comparison with a fixed-radius Euclidean ball gives the asserted cardinality. The same comparison bounds the cardinality of $\mathcal{F}_\alpha$.

Remark 4.28. *Compare with the corresponding statement about covering the sphere with the cubes: the size of the net is κ^n larger.*

Using the net from the Lemma 4.27 and the random rounding construction, we deduce the following.

Lemma 4.29. *Fix $\kappa \geq 1$, $\varepsilon \in (0, 0.1)$, and $\alpha \in \Omega_\kappa$. There exists a finite set $\mathcal{F}_\alpha \subset \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n$ satisfying $\#\mathcal{F}_\alpha \leq (\frac{10\kappa}{\varepsilon})^n$, such that for every $N \times n$ matrix A and every $x \in \mathbb{S}^{n-1}$, there is $y \in \mathcal{F}_\alpha$ with*

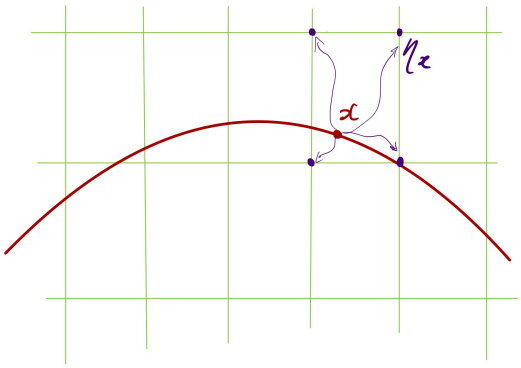
$$|A(x-y)| \leq \frac{\varepsilon}{\sqrt{n}} \sqrt{\sum_{i=1}^n \alpha_i^2 \, |Ae_i|^2}.$$

Keep in mind here that Ae_i are columns of A and $|Ae_i|$ are lengths of columns. Recall that we earlier proved this when $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 1$, in which case we get $\sum_{i=1}^n \alpha_i^2 \, |Ae_i|^2 = \|A\|_{HS}^2$.

Proof. Consider $\mathcal{F}_\alpha$ as in (38). By Lemma 4.27, the size of this net is appropriate, and we have the covering

$$\mathbb{S}^{n-1} \subset \bigcup_{i=1}^m \left(y_i + \frac{\varepsilon}{\sqrt{n}} P_\alpha \right).$$

Fix $x \in \mathbb{S}^{n-1}$ and consider a box P from this covering which contains x. Construct an "α-associated random rounding" η_x, that is, a random vector η_x which takes values in the vertices of P, such that the coordinates of η_x are independent and $\mathbb{E}\,\eta_x = x$.



Then, considering the expected value with respect to the randomness of the random rounding, we get, as before,

$$\begin{aligned} \mathbb{E}\,|A(\eta_x - x)|^2 &= \mathbb{E} \sum_{i=1}^N \langle A^T e_i, \eta_x - x \rangle^2 \\ &\leq \frac{\varepsilon^2}{n} \sum_{i=1}^n \alpha_i^2 \cdot |Ae_i|^2, \end{aligned}$$

by the same reasoning as in the proof of Theorem 2.23, in the equation (8). Consequently, at least one realization has squared error no larger than its expectation. Thus, for every

$x \in \mathbb{S}^{n-1}$, there exists $y \in \mathcal{F}_\alpha$ such that

$$|A(x-y)| \leq \frac{\varepsilon}{\sqrt{n}} \sqrt{\sum_{i=1}^{n} \alpha_i^2 \cdot |Ae_i|^2}.$$

Note that we can indeed guarantee that $y \in \mathcal{F}_\alpha$ because we know that y takes values in the vertices of one of the boxes forming the net $\mathcal{F}_\alpha$. □

This inspires us to formulate the following definition:

Definition 4.30 (A proxy for the Hilbert Schmidt Norm)**.** Let A be an $N \times n$ matrix. We define a functional that will serve as a proxy for the squared Hilbert-Schmidt norm by

$$\mathcal{B}_\kappa(A) := \min_{\alpha \in \Omega_\kappa} \sum_{i=1}^{n} \alpha_i^2 \cdot |Ae_i|^2.$$

Note that $\mathcal{B}_\kappa(A) \leq \|A\|_{HS}^2 = \sum_{i=1}^n |Ae_i|^2$. Therefore, using Lemma 4.29 we can extend Theorem 2.23 as follows:

Corollary 4.31. *Fix $\kappa \geq 1$ and $\varepsilon \in (0, 0.1)$. Let A be an $N \times n$ matrix. Then there exists a net $\mathcal{F} \subset \frac{3}{2}B_2^n \setminus \frac{1}{2}B_2^n$ such that for all $x \in \mathbb{S}^{n-1}$ there exists $y \in \mathcal{F}$ with*

$$|A(x-y)| \leq \frac{\varepsilon}{\sqrt{n}} \sqrt{\mathcal{B}_\kappa(A)},$$

such that $\#\mathcal{F} \leq \left(\frac{10\kappa}{\varepsilon}\right)^n$.

However, Corollary 4.31 alone is not suitable for the union-bound argument we need. Indeed, the value of the vector α which gives the minimum in $\mathcal{B}_\kappa(A)$ depends on the matrix A, and therefore the net in Corollary 4.31 depends on the matrix. In order to apply this to study random matrices, we would like to consider the union of all the nets associated with the admissible boxes in Ω_κ, but there are infinitely many of them. However, it turns out that we can discretize the set Ω_κ, and only consider finitely many nets which would "represent" all the possible nets. We will then be able to construct our desired net as a union of those representative nets.

Step 2: "Nets on nets " Next, we would like to switch the quantifiers in the previous statement: in place of the net that depends on the matrix, we need to have a fixed net, which serves all matrices. For that purpose we shall consider a net on the set of admissible nets.

Lemma 4.32 (Nets on nets)**.** *For every $\kappa > 1$ and $\mu > 0$, there is a finite collection $\mathcal{F} \subset \Omega_{\kappa^{1+\mu}}$ with*

$$\#\mathcal{F} \leq \binom{n + \lfloor n(1+1/\mu) \rfloor}{n} \leq \left(e(2+1/\mu)\right)^n, \tag{39}$$

such that every $\alpha \in \Omega_\kappa$ admits $\beta \in \mathcal{F}$ satisfying $\beta_i \le \alpha_i$ for all i. In particular, for every $N \times n$ matrix A,

$$\mathcal{B}_\kappa(A) \ge \min_{\beta \in \mathcal{F}} \sum_{i=1}^n \beta_i^2 |Ae_i|^2.$$

Proof. Use logarithmic coordinates

$$t_i = \frac{\log(1/\alpha_i)}{\log \kappa} \ge 0.$$

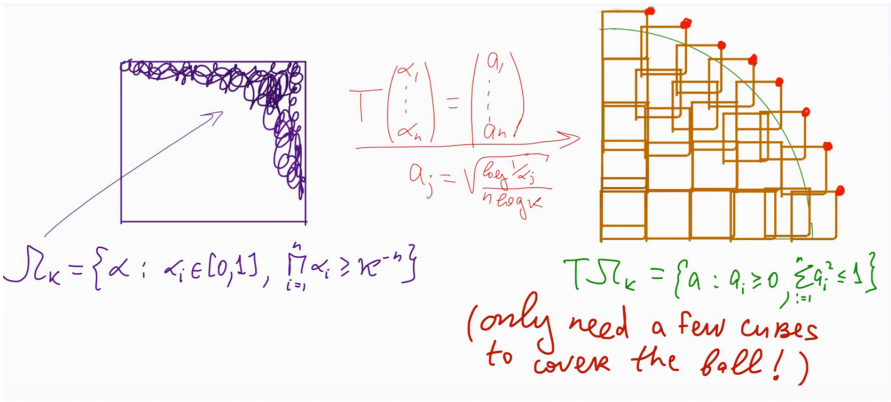
The condition $\alpha \in \Omega_\kappa$ is exactly $\sum_i t_i \le n$. Round each coordinate upward by setting

$$k_i = \left\lceil \frac{t_i}{\mu} \right\rceil, \qquad \beta_i = \kappa^{-\mu k_i}.$$

Then $\beta_i \le \alpha_i$, and

$$\sum_i k_i \le n + n/\mu, \qquad \prod_i \beta_i = \kappa^{-\mu \sum_i k_i} \ge \kappa^{-(1+\mu)n}.$$

Take $\mathcal{F}$ to consist of the vectors $(\kappa^{-\mu k_1}, \dots, \kappa^{-\mu k_n})$ for nonnegative integers k_i with $\sum_i k_i \le \lfloor n(1+1/\mu) \rfloor$. Counting these integer tuples gives the binomial coefficient in (39); the bound $\binom{m}{n} \le (em/n)^n$ gives the second inequality. The last assertion follows by applying the coordinatewise domination to a minimizing α. □



Finally, we deduce the result about a net which serves all deterministic matrices.

Theorem 4.33 (A net for deterministic matrices)**.** *Fix $n \in \mathbb{N}$, $\varepsilon \in (0, 1/10)$ and $\kappa \ge 1$, and let $\mathcal{S} \subset \mathbb{S}^{n-1}$. There is a deterministic net*

$$\mathcal{N} \subset \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n, \qquad \#\mathcal{N} \le \left(\frac{C\kappa(1+\log\kappa)}{\varepsilon}\right)^n,$$

such that, for every $N \times n$ matrix A and every $x \in \mathcal{S}$, some $y \in \mathcal{N}$ satisfies

$$|A(x-y)| \le \frac{\varepsilon}{\sqrt{n}} \sqrt{\mathcal{B}_\kappa(A)}.$$

Proof. For $\kappa > 1$, apply Lemma 4.32 with $\mu = 1/\log\kappa$. It gives a family $\{\beta_1, \dots, \beta_m\} \subset \Omega_{e\kappa}$, with

$$m \le \big(e(2 + \log\kappa)\big)^n,$$

that dominates every admissible α from below, coordinate by coordinate. For each β_i, use Lemma 4.29 to form a net $\mathcal{F}_i$ with at most $(10e\kappa/\varepsilon)^n$ points. For every $x \in \mathbb{S}^{n-1}$, it contains y such that

$$|A(x-y)| \le \frac{\varepsilon}{\sqrt{n}} \left(\sum_{j=1}^n (\beta_i)_j^2 |Ae_j|^2 \right)^{1/2}.$$

Take $\mathcal{N} = \bigcup_i \mathcal{F}_i$ and choose the index i using Lemma 4.32. This proves both assertions. If $\kappa = 1$, then $\Omega_1 = \{(1, \dots, 1)\}$, so Lemma 4.29 applies directly. □

Note that Theorem 4.33 improves upon Theorem 2.23 since $\mathcal{B}_\kappa(A) \le \|A\|_{HS}^2$. It turns out that when the matrix A is random and it has independent columns, this improvement is really crucial, since $\mathcal{B}_\kappa(A)$ has strong large deviation properties, whereas Markov's inequality alone only controls $\|A\|_{HS}^2$ by a fixed multiple of its mean with constant probability. We explore these strong properties in our next step.

Step 3: Large deviation of $\mathcal{B}_\kappa$. Note that if $y_1, \dots, y_n > 0$ are fixed and we constrain $a_i > 0$ and $\prod_{i=1}^n a_i = C_1$ for some constant $C_1 > 0$, then the quantity $\sum_i a_i y_i$ is minimized precisely when $a_i = \frac{C_2}{y_i}$ for all i and some constant $C_2 > 0$. This inspires our next proof.

Lemma 4.34. *Let A be an $N \times n$ random matrix with independent columns and finite $\mathbb{E}\,\|A\|_{HS}^2$. For every $\kappa \ge 1$,*

$$\mathbb{P}\{\mathcal{B}_\kappa(A) > 10\,\mathbb{E}\,\|A\|_{HS}^2\} \le \left(\frac{6}{5\kappa^2}\right)^n.$$

Proof. Set $Y_i = |Ae_i|$, $m_i = \mathbb{E}\,Y_i^2$, and

$$a_i = \sqrt{\min\{1, 5m_i/Y_i^2\}}.$$

If $Y_i = 0$, set $a_i = 1$; a column with $m_i = 0$ is zero almost surely and is treated in the same way. In every case,

$$\sum_i a_i^2 Y_i^2 \le 5 \sum_i m_i = 5\,\mathbb{E}\,\|A\|_{HS}^2.$$

Thus, if $a = (a_1, \dots, a_n)$ belongs to Ω_κ, the event on the left-hand side cannot occur. Moreover,

$$\mathbb{P}\{a \notin \Omega_\kappa\} = \mathbb{P}\left\{ \prod_i a_i^{-2} > \kappa^{2n} \right\} \le \kappa^{-2n}\,\mathbb{E} \prod_i a_i^{-2}.$$

The variables a_i are independent, since each depends only on the i-th column. When $m_i > 0$,

$$\mathbb{E}\,a_i^{-2} = \mathbb{E}\max\{1, Y_i^2/(5m_i)\} \le 1 + \frac{1}{5} = \frac{6}{5};$$

for a zero column the expectation equals 1. Multiplying these bounds finishes the proof. Notice that keeping the maximum with 1 is essential when estimating the product. □

Proof of Theorem 4.25. Apply Theorem 4.33 with $\kappa = 20$ and then Lemma 4.34. Since $(6/(5\cdot 20^2))^n \le e^{-5n}$, the good event has the required probability. On this event the approximation error is at most $\sqrt{10}\varepsilon\sqrt{\mathbb{E}\,\|A\|_{HS}^2}/\sqrt{n}$, and all fixed numerical factors can be absorbed into the absolute constant C. □

Remark 4.35 (Important remark)**.** *For a fixed small ε, the construction above gives a net of size C^n, with an absolute base that can be considerably larger than that of a usual Euclidean ε-net. A construction retaining all vertices of a single nondegenerate coordinate box already contains 2^n points. Often this extra size does not matter.*

However, suppose $S \subset \mathbb{S}^{n-1}$ is a "small set" which can be covered by m $\varepsilon-$balls, where $m \le 1.01^n$. So S is in this sense a lot smaller than the whole sphere (whose covering number at a fixed small radius has a much larger exponential rate). Then S can also be covered by parallelepipeds and one can construct a net on S with properties like the ones in Theorem 4.25. And furthermore, for any $\gamma > 0$ one can have a net $\mathcal{F}$ near S of size $\#\mathcal{F} \le (1+\gamma)^n m$ (recall that m is the number of $\varepsilon-$balls covering S), and such that, with probability at least $1 - e^{-c(\gamma)n}$, simultaneously for every $x \in S$ there is $y \in \mathcal{F}$ satisfying

$$|A(x-y)| \le \frac{C(\gamma)\varepsilon\sqrt{\mathbb{E}\,\|A\|_{HS}^2}}{\sqrt{n}}.$$

Here ε is sufficiently small depending on γ, and the assumptions on A are the same as in Theorem 4.25. For the full statement and the details, see Livshyts [26].

At first glance, this may seem surprising: how could one guarantee a net made out of vertices of some boxes of size less than, say 1.2^n, if there are 2^n vertices already for one box? The trick is that a lot of the vertices could be "dismissed" from this construction. This requires some further ideas and technicalities which we leave beyond the scope of this course. See the details in [26], particularly the refined net theorem and its corollaries.

4.6 The smallest singular value of square random matrices

The above concentration results for the smallest singular value of A rely upon taking the number of rows of A to be sufficiently large relative to the number of columns. Having $N \ge Cn$ allowed the term

$$\sup_{x\in\mathbb{S}^{n-1}} \mathbb{P}(|Ax| \le \clubsuit) \le e^{-cN}$$

compensate for the $e^{C_1 n}$ factor coming from the union bound applied over the ε-net. The standard ε-net argument no longer works if A is a square matrix, even if one imposes additional strong assumptions on the matrix – it doesn't even work for Gaussian random matrices. However, when an idea does not work, it might be a good strategy to find a way to still apply it to make at least some progress.

In what follows we describe a foundational idea of Rudelson and Vershynin about a decomposition of the sphere. Instead of using an ε-net to cover the whole sphere $\mathbb{S}^{n-1}$, we can construct an ε-net of prescribed size e^{cn} for a "small but well spread" subset $S \subseteq \mathbb{S}^{n-1}$. "Small" in this context means that the set has a small covering number, and "well spread" means that its complement is in some sense predictable and controllable using different ideas. The smallest singular value can then be analyzed by noticing that

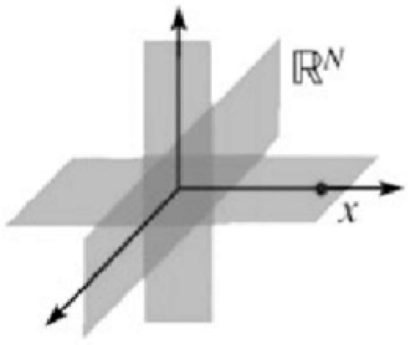
$$\mathbb{P}\left(\inf_{x\in\mathbb{S}^{n-1}} |Ax| \le \clubsuit\right) \le \mathbb{P}\left(\inf_{x\in S} |Ax| \le \clubsuit\right) + \mathbb{P}\left(\inf_{x\in\mathbb{S}^{n-1}\setminus S} |Ax| \le \clubsuit\right),$$

and considering the infima over S and over its complement in $\mathbb{S}^{n-1}$ separately.

Definition 4.36. For $\delta \in (0,1)$ define

$$\mathrm{Sparse}(\delta) := \{x \in \mathbb{R}^n : \#\{i : x_i \neq 0\} \le \lfloor \delta n \rfloor\}.$$

Thus $\mathrm{Sparse}(\delta)$ is the union of coordinate subspaces of dimension at most $\lfloor \delta n \rfloor$. When discussing coverings of the sphere, we use $\mathrm{Sparse}(\delta) \cap \mathbb{S}^{n-1}$. Allowing sparse vectors of arbitrary length will also let us use coordinate projections below.



The following Lemma is a crucial fact about sparse vectors: the set of unit sparse vectors can be covered by a relatively small number of ε-balls.

Lemma 4.37. *For $\varepsilon \in (0,1)$ and $\delta \in (0,1/2)$,*

$$\mathrm{Sparse}(\delta) \cap \mathbb{S}^{n-1} \subset \bigcup_{i=1}^{m} (y_i + \varepsilon \mathbf{B}_2^n), \qquad m \le \left(\frac{3e}{\delta\varepsilon}\right)^{\delta n}.$$

Proof. Write $k = \lfloor \delta n \rfloor$. If $k = 0$, the set on the left is empty. Otherwise, there are $\binom{n}{k}$ coordinate subspaces of dimension k, and the unit sphere in each admits an ε-net of size at most $(3/\varepsilon)^k$. Therefore,

$$m \le \binom{n}{k}\left(\frac{3}{\varepsilon}\right)^k \le \left(\frac{en}{k}\right)^k \left(\frac{3}{\varepsilon}\right)^k \le \left(\frac{3e}{\delta\varepsilon}\right)^{\delta n}.$$

The binomial estimate follows, for example, from $k! \ge (k/e)^k$. In the last step we used that $r \mapsto r\log(en/r)$ is increasing for $0 < r \le n$. Thus, for fixed ε, the exponential rate of the covering number tends to zero as $\delta \downarrow 0$. □

4.6.1 Rudelson–Vershynin decomposition of the sphere

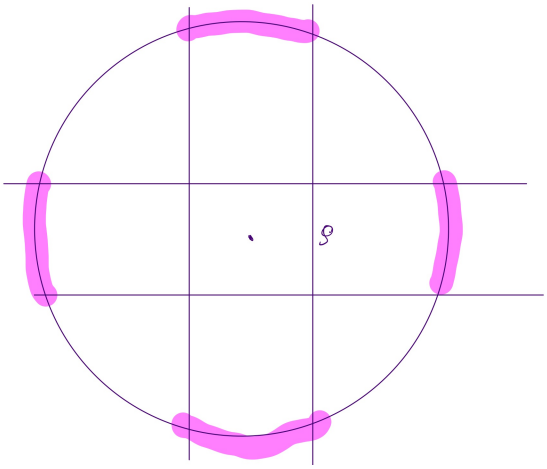
As we discussed earlier, we want to decompose the unit sphere into two disjoint subsets, one of which would have a small covering. Sparse vectors are a good candidate for this first set. However, that is not enough: the set of sparse vectors has measure zero, and removing it does not achieve much... So instead we will do a net argument on the set of vectors which are *close to sparse vectors*. These are called *compressible vectors*. Here is the formal

Definition 4.38 (the Rudelson-Vershynin decomposition of the sphere)**.** For $\delta, \rho \in (0,1)$, define the *compressible vectors* to be the set

$$\begin{aligned}\mathrm{Comp}(\delta,\rho) &= \{x \in \mathbb{S}^{n-1} : \mathrm{dist}(x, \mathrm{Sparse}(\delta)) \le \rho\} \\ &= \{x \in \mathbb{S}^{n-1} : \exists\, y \in \mathrm{Sparse}(\delta) \text{ s.t. } |x-y| \le \rho\},\end{aligned}$$

and the *incompressible vectors* to be their complement

$$\mathrm{Incomp}(\delta,\rho) = \mathbb{S}^{n-1} \setminus \mathrm{Comp}(\delta,\rho).$$



Lemma 4.37 implies:

Lemma 4.39. *For every $c > 0$ and $\varepsilon \in (0,1)$, there are $\delta, \rho \in (0, 1/2)$ such that*

$$\mathrm{Comp}(\delta,\rho) \subset \bigcup_{i=1}^{m} (y_i + \varepsilon \mathbf{B}_2^n), \qquad m \le e^{cn}.$$

Proof. Take $\rho = \varepsilon/4$. If $x \in \mathrm{Comp}(\delta,\rho)$ and $y \in \mathrm{Sparse}(\delta)$ satisfies $|x-y| \le \rho$, then $y \neq 0$ and $|x - y/|y|| \le 2\rho$. Cover $\mathrm{Sparse}(\delta) \cap \mathbb{S}^{n-1}$ by balls of radius $\varepsilon/2$ using Lemma 4.37, and choose δ so small that $\delta \log(6e/(\delta\varepsilon)) \le c$. □

The following claim follows from this lemma and the refined net construction in Remark 4.35.

Claim 4.40. *For every $c > 0$ and $\eta > 0$, there are $\delta, \rho \in (0, 1/2)$, $c_0 > 0$, and a deterministic finite set*

$$\mathcal{N} \subset \frac{3}{2}\mathbf{B}_2^n \setminus \frac{1}{2}\mathbf{B}_2^n, \qquad \#\mathcal{N} \leq e^{cn},$$

such that for any random $N \times n$ matrix A with independent columns and finite $\mathbb{E}\,\|A\|_{HS}^2$, the following holds with probability at least $1 - e^{-c_0 n}$: every $x \in \mathrm{Comp}(\delta, \rho)$ has $y \in \mathcal{N}$ satisfying

$$|A(x - y)| \leq \frac{\eta}{\sqrt{n}} \sqrt{\mathbb{E}\,\|A\|_{HS}^2}.$$

The parameters depend only on c, η. In particular, the approximation coefficient can be chosen as small as needed, at the cost of changing δ, ρ.

The Claim above will be useful for us when we estimate the infimum of $|Ax|$ over the set of compressible vectors. But for the time being, let us discuss how the smallest singular value of square random matrices actually behaves, and state some relevant results.

4.6.2 Survey of results regarding the smallest singular value of square random matrices

Theorem 4.41 (Rudelson-Vershynin)**.** *Let A be an $n \times n$ random matrix with entries that are iid, K-sub-Gaussian, mean-zero, and variance-one. For all $\varepsilon > 0$ one has*

$$\mathbb{P}\left\{\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right\} \leq C\varepsilon + e^{-c_1 n},$$

where C and c_1 are positive constants depending only on K.

In the analogous setting for tall matrices $N \geq Cn$, we established that $\mathbb{E}\,\sigma_n(A) \geq \widetilde{c}\sqrt{N}$, which tends to infinity. Theorem 4.41 states something quite different for square matrices: the smallest singular value turns out to be of order $\frac{1}{\sqrt{n}}$ on average, and in fact this estimate is sharp (up to constants depending on K), see Rudelson, Vershynin [37], Tatarko [46]. Let us consider two key examples which demonstrate the sharpness also of the probability estimate in Theorem 4.41.

Example 4.42. *Suppose the entries of A are i.i.d. distributed as standard Gaussian random variables. Szarek [45] and Edelman [8] proved that for all $\varepsilon > 0$ one has*

$$\mathbb{P}\left\{\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right\} \leq \varepsilon.$$

Note that this implies $\mathbb{P}\{\sigma_n(A) = 0\} = 0$, and this phenomenon would more generally be true for any square random matrix with independent, non-atomic entries, even without a mean-zero assumption, see Tikhomirov [53].

Example 4.43. *Suppose the entries a_{ij} are i.i.d. symmetric Bernoulli random variables with parameter $1/2$, i.e. $a_{ij} \sim \mathrm{Unif}\{-1, 1\}$. What can be said about the invertibility of A? One has for instance*

$$\begin{aligned}\mathbb{P}\{\sigma_n(A) = 0\} &\geq \mathbb{P}\left\{\textit{two rows are the same or two columns are the same}\right\} \\ &\geq (1 - o(1))n^2 2^{-n}.\end{aligned}$$

A conjecture of Erdős stated that this bound is sharp up to a polynomial error. Tikhomirov [54] determined the exponential rate by proving

$$\mathbb{P}\{\sigma_n(A) = 0\} = (1/2 + o(1))^n.$$

Thus an exponentially small additive error is unavoidable in a theorem that includes Bernoulli matrices; this statement does not assert an exact universal rate for all discrete distributions.

The Gaussian and Bernoulli examples together show that the probability estimate in Theorem 4.41 is natural.

Following the result from Theorem 4.41, Rebrova and Tikhomirov were able to relax the sub-Gaussian assumption, instead considering uniformly anti-concentrated random variables.

Theorem 4.44 (Rebrova-Tikhomirov [33])**.** *Let A be an $n \times n$ random matrix whose entries a_{ij} are zero-mean, unit variance, and i.i.d. If a_{ij} are uniformly anti-concentrated (UAC),* i.e., $\sup_{z \in \mathbb{R}} \mathbb{P}(|a_{ij} - z| \leq a) \leq b$ *for fixed $a > 0$, $b \in (0, 1)$, then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq c\varepsilon + e^{-c_1 n},$$

where c, c_1 are constants that depend on a and b.

The proof technique for Theorem 4.44 uses a more clever ε-net argument than Theorem 4.41, which depends on the norm of the matrix. Their proof uses the moment assumptions and uniform anti-concentration, while the net construction avoids bounding the full operator norm. A further generalization of the result investigates removing the assumption that the entries are all identically distributed, zero-mean, and unit variance.

Theorem 4.45 (Livshyts [26])**.** *Let A be an $n \times n$ random matrix whose entries a_{ij} are independent. If $\mathbb{E}\,\|A\|_{HS}^2 \leq Kn^2$, and a_{ij} are uniformly anti-concentrated (UAC),* i.e., $\sup_{z \in \mathbb{R}} \mathbb{P}(|a_{ij} - z| \leq a) \leq b$ *for fixed $a > 0$, $b \in (0, 1)$, then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq c\varepsilon + \frac{c_1}{\sqrt{n}}, \tag{40}$$

where c, c_1 are constants that depend only on a, b, K. Moreover, if the rows of A are i.i.d., then

$$\mathbb{P}\left(\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq c\varepsilon + 2e^{-c_1 n}, \tag{41}$$

where c, c_1 are constants that depend only on a, b, K.

Remark 4.46. *The polynomial bound from* (40)*, while easier to prove, is a strictly weaker bound than the exponential bound of* (41)*, which recovers the bound from Theorem 4.44 for the i.i.d. case. Recovering the exponential bound* (41) *requires an additional tool called the "Least Common Denominator" (LCD) of vectors. Both error terms tend to zero with* n*, but their different rates matter when one needs quantitative probabilities. The proof of the polynomial bound* (40)*, is worked out in the next subsection.*

In subsequent work, Livshyts, Tikhomirov, and Vershynin were able to prove the same exponential bound without the i.i.d. row assumption.

Theorem 4.47 (Livshyts, Tikhomirov, Vershynin [27])**.** *Let* A *be an* $n \times n$ *random matrix whose entries* a_{ij} *are independent. If* $\mathbb{E}\,\|A\|_{HS}^2 \leq Kn^2$*, and* a_{ij} *are uniformly anti-concentrated (UAC),* i.e.*,* $\sup_{z\in\mathbb{R}} \mathbb{P}(|a_{ij} - z| \leq a) \leq b$ *for fixed* $a > 0$*,* $b \in (0, 1)$*, then for every* $\varepsilon > 0$*,*

$$\mathbb{P}\left(\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq c\varepsilon + 2e^{-c_1 n},$$

where c, c_1 *are constants that depend only on* a, b, K*.*

The proof of Theorem 4.47 requires the use of the "randomized LCD", similar to the LCD used in the proof of Theorem 4.45.

Arbitrary aspect ratios We briefly state some results for matrices of arbitrary aspect ratios. Consider the random matrix A, of dimension $N \times n$, with i.i.d., zero mean, and unit variance entries a_{ij}. Consider the case $N \geq n$, but possibly not $N >> n$.

Theorem 4.48 (Rudelson, Vershynin [38])**.** *If* a_{ij} *are sub-Gaussian, then for every* $\varepsilon > 0$*,*

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon(\sqrt{N+1} - \sqrt{n})) \leq (c\varepsilon)^{N-n+1} + e^{-c_1 n},$$

for positive constants c, c_1 *depending only on the sub-Gaussian norm.*

Theorem 4.49 (Livshyts [26])**.** *Assume the entries are i.i.d., with mean zero and variance one as above. If they satisfy UAC with parameters* a, b*, then for every* $0 < \varepsilon < 1/2$*,*

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon(\sqrt{N+1} - \sqrt{n})) \leq \left(c\varepsilon \log \frac{1}{\varepsilon}\right)^{N-n+1} + e^{-c_1 n},$$

for constants c, c_1 *dependent on* a *and* b*.*

4.7 Proof of Theorem 4.45 Part 1

Recall the statement of the theorem. Let A be an $n \times n$ random matrix whose entries a_{ij} are independent. If $\mathbb{E}\,\|A\|_{HS}^2 \le Kn^2$, and a_{ij} are uniformly anti-concentrated (UAC), *i.e.*, $\sup_{z\in\mathbb{R}} \mathbb{P}(|a_{ij} - z| \le a) \le b$ for fixed $a > 0$, $b \in (0,1)$, then for every $\varepsilon > 0$,

$$\mathbb{P}\left(\sigma_n(A) \le \frac{\varepsilon}{\sqrt{n}}\right) \le c\varepsilon + \frac{c_1}{\sqrt{n}},$$

where c, c_1 are constants that depend only on a, b, K.

To prove this result, we will need to use additional tools to separately handle the compressible and incompressible vectors from the Rudelson-Vershynin decomposition of the sphere from Definition 4.38. Recall the following definitions of sparse vectors, compressible vectors (close to sparse), and incompressible vectors (far from sparse). For $\delta, \rho \in (0,1)$,

$$\begin{aligned}
\mathrm{Sparse}(\delta) &= \{x \in \mathbb{R}^n : \#\{i : x_i \neq 0\} \le \lfloor \delta n \rfloor\}, \\
\mathrm{Comp}(\delta,\rho) &= \{x \in \mathbb{S}^{n-1} : \exists\, y \in \mathrm{Sparse}(\delta) \text{ s.t. } |x - y| \le \rho\}, \\
\mathrm{Incomp}(\delta,\rho) &= \mathbb{S}^{n-1} \setminus \mathrm{Comp}(\delta,\rho).
\end{aligned}$$

4.7.1 Compressible Vectors

After taking the Rudelson-Vershynin decomposition of the sphere, we first consider the compressible vectors in $\mathrm{Comp}(\delta,\rho)$. The following Lemma bounds the behavior of the compressible vectors.

Lemma 4.50 (Compressible vectors)**.** *Let A be an $n \times n$ random matrix with independent entries satisfying*

$$\sup_{z\in\mathbb{R}} \mathbb{P}(|a_{ij} - z| \le a) \le b < 1, \qquad \mathbb{E}\,\|A\|_{HS}^2 \le Kn^2.$$

There are $\delta, \rho \in (0, 1/2)$ and $c_0, c_1 > 0$, depending only on a, b, K, such that

$$\mathbb{P}\left\{\inf_{x\in\mathrm{Comp}(\delta,\rho)} |Ax| \le c_0\sqrt{n}\right\} \le 2e^{-c_1 n}.$$

Proof. By Corollary 4.22, there are $d_0, d_1 > 0$, depending only on a, b, such that

$$\sup_{|u|=1} \mathbb{P}\{|Au| \le d_0\sqrt{n}\} \le e^{-d_1 n}.$$

Apply Claim 4.40 with entropy exponent $d_1/2$ and approximation coefficient η chosen so that $\eta\sqrt{K} \le d_0/4$. It gives a deterministic net $\mathcal{N}$ of size at most $e^{d_1 n/2}$ and a good event of probability at least $1 - e^{-d_2 n}$ on which

$$\forall x \in \mathrm{Comp}(\delta,\rho) \quad \exists y \in \mathcal{N} : \quad |A(x-y)| \le \eta \frac{\sqrt{\mathbb{E}\,\|A\|_{HS}^2}}{\sqrt{n}} \le \frac{d_0}{4}\sqrt{n}.$$

Take $c_0 = d_0/4$. On the good event, if some compressible x has $|Ax| \le c_0\sqrt{n}$, its net point satisfies $|Ay| \le (d_0/2)\sqrt{n}$. Since every $y \in \mathcal{N}$ has $|y| \ge 1/2$, the pointwise estimate and a union bound give

$$\begin{aligned}\mathbb{P}\left\{\inf_{x\in\mathrm{Comp}(\delta,\rho)} |Ax| \le c_0\sqrt{n}\right\} &\le e^{-d_2 n} + \sum_{y\in\mathcal{N}} \mathbb{P}\{|Ay| \le (d_0/2)\sqrt{n}\} \\ &\le e^{-d_2 n} + e^{d_1 n/2} e^{-d_1 n} \le 2e^{-c_1 n}.\end{aligned}$$

□

Remark 4.51. *Independent columns were used in constructing the net, the Hilbert–Schmidt moment bound controlled the approximation error, and independent rows together with uniform anti-concentration gave the pointwise small-ball estimate. The same proof works for an $(n-1)\times n$ matrix, for $n \ge 2$, with the same type of assumptions and adjusted constants: $n - 1 \ge n/2$, so the pointwise estimate still decays exponentially in n.*

4.7.2 Incompressible Vectors

While the result for the compressible vectors followed immediately from previous results, we will need to use different tools to handle the case of incompressible vectors in $\mathrm{Incomp}(\delta,\rho)$. The following Example demonstrates some crucial behavior of incompressible vectors which will be useful for bounding their behavior.

Example 4.52 (An incompressible vector)**.** *Consider $x = (1/\sqrt{n},\dots,1/\sqrt{n}) \in \mathbb{S}^{n-1}$. For every $y \in \mathrm{Sparse}(\delta)$, at least $n - \lfloor \delta n \rfloor$ coordinates of y vanish. Therefore*

$$|x - y|^2 \ge \frac{n - \lfloor \delta n \rfloor}{n} \ge 1 - \delta.$$

Thus $x \in \mathrm{Incomp}(\delta,\rho)$ whenever $\rho < \sqrt{1-\delta}$. This vector is spread over all coordinates, each of magnitude $1/\sqrt{n}$.

Example 4.52 provides some intuition for how an incompressible vector behaves. In what follows, we will try to find some characterization of this behavior, where incompressible vectors will similarly have a substantial subset of coordinates where $|x_k| \sim \frac{1}{\sqrt{n}}$.

Lemma 4.53 (Incompressible vectors are spread)**.** *For every $x \in \mathrm{Incomp}(\delta,\rho)$, for $\delta,\rho \in (0,1)$, there exists a subset of indices $\sigma \subset \{1,\dots,n\}$ with $\#\sigma \ge \frac{1}{2}\rho^2\delta n$ such that for every $k \in \sigma$,*

$$\frac{\rho}{\sqrt{2n}} \le |x_k| \le \frac{1}{\sqrt{\delta n}}.$$

Proof. Let $x \in \mathrm{Incomp}(\delta,\rho)$, and consider $\sigma_1,\sigma_2 \subset \{1,\dots,n\}$ such that

$$\sigma_1 := \left\{k : |x_k| \le \frac{1}{\sqrt{\delta n}}\right\}, \quad \sigma_2 := \left\{k : |x_k| \ge \frac{\rho}{\sqrt{2n}}\right\}.$$

We would like to show that the cardinality of $\sigma = \sigma_1 \cap \sigma_2$ is controlled, *i.e.*, that $\#\sigma \geq cn$ for some c. First, since $x \in \mathbb{S}^{n-1}$, $|x| = 1$, and thus,

$$1 = |x|^2 = \sum_{k=1}^{n} |x_k|^2 \geq \sum_{k \in \sigma_1^{\complement}} |x_k|^2 \geq \sum_{k \in \sigma_1^{\complement}} \left| \frac{1}{\sqrt{\delta n}} \right|^2 \geq \#\sigma_1^{\complement} \frac{1}{\delta n}.$$

Thus, $\#\sigma_1^{\complement} \leq \delta n$, so $\#\sigma_1 \geq n - \delta n$. Next, consider the following projection operator P_σ,

$$P_\sigma(x) = (y_1, y_2, \ldots, y_n), \quad y_k = \begin{cases} x_k & k \in \sigma \\ 0 & k \notin \sigma \end{cases}.$$

Let $y = P_{\sigma_1^{\complement}}(x)$. The support of y has at most $\lfloor \delta n \rfloor$ elements, so $y \in \text{Sparse}(\delta)$. Since $x \in \text{Incomp}(\delta, \rho)$, we have $|x - y| > \rho$. Note that

$$|P_{\sigma_2^{\complement}}(x)|^2 = \sum_{k : |x_k| < \frac{\rho}{\sqrt{2n}}} |x_k|^2 \leq n \left| \frac{\rho}{\sqrt{2n}} \right|^2 = \frac{\rho^2}{2}, \text{ and } |P_{\sigma_1}(x)|^2 = |x - y|^2 \geq \rho^2,$$

which implies that

$$|P_\sigma(x)|^2 \geq |P_{\sigma_1}(x)|^2 - |P_{\sigma_2^{\complement}}(x)|^2 \geq \frac{\rho^2}{2}.$$

On the other hand,

$$|P_\sigma(x)|^2 \leq \#\sigma \max_{k \in \sigma} |x_k|^2 \leq \frac{1}{\delta n} \#\sigma,$$

and therefore, $\#\sigma \geq \delta n \frac{\rho^2}{2}$. □

Lemma 4.54 (Invertibility via distance)**.** *Let A be any random $N \times n$ matrix, let $X_j = Ae_j$, and put $H_j = \text{span}\{X_i : i \neq j\}$. For $\delta, \rho \in (0, 1/2)$ and $\varepsilon > 0$,*

$$\mathbb{P}\left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} |Ax| \leq \frac{\rho \varepsilon}{\sqrt{n}} \right\} \leq \frac{1}{\delta n} \sum_{j=1}^{n} \mathbb{P}\{\text{dist}(X_j, H_j) \leq \varepsilon\}.$$

No independence assumption is needed for this lemma.

Proof. For every $x \in \mathbb{S}^{n-1}$ and every k, all terms $x_j X_j$ with $j \neq k$ belong to H_k. Hence

$$|Ax| \geq \text{dist}(Ax, H_k) = |x_k| \, \text{dist}(X_k, H_k). \tag{42}$$

Write $p_k = \mathbb{P}\{\text{dist}(X_k, H_k) \leq \varepsilon\}$, and let $\mathcal{U}$ be the event that

$$\sigma_1 := \{k : \text{dist}(X_k, H_k) > \varepsilon\} \quad \text{has more than } (1 - \delta)n \text{ elements.}$$

Markov's inequality gives

$$\mathbb{P}(\mathcal{U}^c) \le \frac{1}{\delta n}\,\mathbb{E}\,\#\{k : \operatorname{dist}(X_k, H_k) \le \varepsilon\} = \frac{1}{\delta n}\sum_{k=1}^{n} p_k. \tag{43}$$

For $x \in \operatorname{Incomp}(\delta, \rho)$, set

$$\sigma_2(x) = \{k : |x_k| \ge \rho/\sqrt{n}\}.$$

This set has more than δn elements. Otherwise its coordinate projection $y = P_{\sigma_2(x)}x$ would be sparse and would satisfy

$$|x - y|^2 = \sum_{k \notin \sigma_2(x)} |x_k|^2 \le \rho^2,$$

contradicting incompressibility. On $\mathcal{U}$, the two sets σ_1 and $\sigma_2(x)$ therefore intersect. By (42),

$$|Ax| \ge \frac{\rho}{\sqrt{n}} \min_{k \in \sigma_1} \operatorname{dist}(X_k, H_k) > \frac{\rho\varepsilon}{\sqrt{n}}.$$

The minimum here does not depend on x, so the same strict inequality holds for the infimum over all incompressible vectors. Thus the event on the left of the lemma is contained in $\mathcal{U}^c$, and (43) finishes the proof. □

Lemma 4.54 reduces the problem from dealing with the incompressible vectors to finding an estimate of the following form

$$\mathbb{P}(\operatorname{dist}(X, H) \le \varepsilon) \le c\varepsilon + \frac{c_1}{\sqrt{n}},$$

where X is a random vector with independent UAC entries, and H is the span of $n-1$ independent random vectors with independent UAC entries.

4.7.3 Distance Theorem

Recall that our goal is to prove the following 'distance' theorem

Theorem 4.55 (Distance theorem)**.** *Let $X \in \mathbb{R}^n$ have independent coordinates, and let B be an independent random $n \times (n-1)$ matrix with independent entries. Assume that all coordinates of X and all entries of B are uniformly anti-concentrated with common parameters a, b, and that $\mathbb{E}\,\|B\|_{HS}^2 \le Kn^2$. Set $H = \operatorname{im} B$. Then*

$$\mathbb{P}\{\operatorname{dist}(X, H) \le \varepsilon\} \le C\varepsilon + \frac{C_1}{\sqrt{n}}, \qquad \varepsilon > 0, \tag{44}$$

where C, C_1 depend only on a, b, K.

Remark 4.56. *The sharper distance theorem in [27] has an exponentially small error. Its application also involves a quantitative second-moment condition on the distinguished column and a step that excludes a small number of exceptional columns. The polynomial estimate above avoids those additional ingredients.*

If H is a hyperplane with unit normal ν, then $\operatorname{dist}(X, H) = |\langle X, \nu\rangle|$. In general H may have smaller dimension; for every unit $\nu \perp H$ we still have

$$\operatorname{dist}(X, H) \geq |\langle X, \nu\rangle|,$$

which is the direction needed for an upper bound on the small-ball probability.

Theorem 4.57 (Rogozin's Theorem)**.** *Let $v = (v_1, \cdots, v_n)$ be a random vector with independent UAC entries, say*

$$\sup_{z\in\mathbb{R}} \mathbb{P}\left(|v_i - z| < a\right) < b. \tag{45}$$

Then for any nonzero $u \in \mathbb{R}^n$ and any $\varepsilon > ca\|u\|_\infty$

$$\sup_{z\in\mathbb{R}} \mathbb{P}\left(|\langle u, v\rangle - z| < \varepsilon\right) \leq \frac{C\varepsilon}{|u|}, \tag{46}$$

where C, c depend only on a and b.

Corollary 4.58. *Fix $c_1 > 0$ and $c_2 \in (0, 1]$. Let $u \in \mathbb{S}^{n-1}$ satisfy*

$$\#\{i : |u_i| \geq c_1/\sqrt{n}\} \geq c_2 n.$$

If v has independent UAC coordinates, then

$$\sup_{z\in\mathbb{R}} \mathbb{P}\{|\langle u, v\rangle - z| \leq \varepsilon\} \leq C_2\varepsilon \qquad \text{for } \varepsilon \geq C_1/\sqrt{n},$$

where C_1, C_2 depend only on the UAC parameters and c_1, c_2. The conclusion also holds for a random unit vector u independent of v, if the displayed spread condition holds almost surely.

Proof. First fix u. Since $|u| = 1$, at most $c_2 n/2$ coordinates can have magnitude greater than $\sqrt{2/c_2}/\sqrt{n}$. Thus

$$\sigma = \left\{ i : \frac{c_1}{\sqrt{n}} \leq |u_i| \leq \frac{\sqrt{2/c_2}}{\sqrt{n}} \right\}$$

has at least $c_2 n/2$ elements, and

$$\sum_{i\in\sigma} u_i^2 \geq c_1^2 c_2/2.$$

Write $\langle u, v\rangle = R + \sum_{i\in\sigma} u_i v_i$, where $R = \sum_{i\notin\sigma} u_i v_i$. Independence and conditioning on R show that adding R cannot increase the supremum of the small-ball probabilities. Apply Rogozin's theorem to the restricted sum: its largest coefficient is at most $\sqrt{2/c_2}/\sqrt{n}$, while its coefficient vector has norm at least $c_1\sqrt{c_2/2}$. This gives the required bound for $\varepsilon \geq C_1/\sqrt{n}$, after increasing C_1 if necessary. Closed intervals follow by slightly enlarging the radius. If u is random and independent of v, condition on u and integrate. □

By monotonicity in ε, the same argument gives the convenient all-scale estimate

$$\sup_{z\in\mathbb{R}} \mathbb{P}\{|\langle u, v\rangle - z| \le \varepsilon\} \le C\varepsilon + \frac{C'}{\sqrt{n}}, \qquad \varepsilon > 0.$$

We apply this with $X = Ae_j$ and with a unit normal ν to $H_j = \operatorname{span}(Ae_i : i \neq j)$, chosen using only the other columns. Then X and ν are independent. It remains to show that ν is incompressible with high probability.

Lemma 4.59 (Random normals are incompressible)**.** *Let B be an $n \times (n-1)$ random matrix with independent UAC entries and $\mathbb{E}\,\|B\|_{HS}^2 \le Kn^2$, and put $H = \operatorname{im} B$. There are $\delta, \rho, c > 0$, depending only on the UAC parameters and K, such that, with probability at least $1 - 2e^{-cn}$, every unit vector $\nu \perp H$ belongs to* $\operatorname{Incomp}(\delta, \rho)$.

Proof. The relation $\nu \perp H$ is equivalent to $B^T\nu = 0$. Therefore,

$$\mathbb{P}\{\exists \nu \in \operatorname{Comp}(\delta,\rho) : \nu \perp H\} \le \mathbb{P}\left\{\inf_{x\in\operatorname{Comp}(\delta,\rho)} |B^T x| = 0\right\} \le 2e^{-cn}.$$

The last step is the $(n-1)\times n$ version of Lemma 4.50 noted after its proof. The case $n = 1$ can be absorbed by increasing the constants in the distance estimate. □

Proof of Theorem 4.55. Choose a unit $\nu \perp H$ as a function of B. It is independent of X. On the good event from Lemma 4.59, Lemma 4.53 gives a fixed positive fraction of coordinates of ν with magnitude at least $c/\sqrt{n}$. Conditional on B, Corollary 4.58 and its all-scale consequence therefore give

$$\begin{aligned}\mathbb{P}\{\operatorname{dist}(X, H) \le \varepsilon\} &\le \mathbb{P}\{|\langle X, \nu\rangle| \le \varepsilon\} \\ &\le 2e^{-cn} + C\varepsilon + \frac{C'}{\sqrt{n}} \le C\varepsilon + \frac{C_1}{\sqrt{n}}.\end{aligned}$$

This proves the distance theorem, including the possible case $\dim H < n - 1$. □

4.7.4 Proof of the first part of Theorem 4.45

Recall that A has independent UAC entries and $\mathbb{E}\,\|A\|_{HS}^2 \le Kn^2$. Let δ, ρ, c_0, c_1 be the constants in Lemma 4.50, and set $H_j = \operatorname{span}(Ae_i : i \neq j)$. For $0 < \varepsilon \le c_0$, the decomposition of the sphere, the compressible-vector bound, and Lemma 4.54 give

$$\begin{aligned}\mathbb{P}\left\{\sigma_n(A) \le \frac{\varepsilon}{\sqrt{n}}\right\} &\le \mathbb{P}\left\{\inf_{x\in\operatorname{Comp}(\delta,\rho)} |Ax| \le \frac{\varepsilon}{\sqrt{n}}\right\} + \mathbb{P}\left\{\inf_{x\in\operatorname{Incomp}(\delta,\rho)} |Ax| \le \frac{\varepsilon}{\sqrt{n}}\right\} \\ &\le 2e^{-c_1 n} + \frac{1}{\delta n}\sum_{j=1}^{n} \mathbb{P}\{\operatorname{dist}(Ae_j, H_j) \le \varepsilon/\rho\}.\end{aligned}$$

For each j, the column Ae_j is independent of the matrix consisting of all the other columns, and that matrix still has expected squared Hilbert–Schmidt norm at most Kn^2. The distance theorem therefore applies uniformly in j. We obtain

$$\begin{aligned}\mathbb{P}\left\{\sigma_n(A) \le \frac{\varepsilon}{\sqrt{n}}\right\} &\le 2e^{-c_1 n} + \frac{1}{\delta}\left(\frac{C\varepsilon}{\rho} + \frac{C_1}{\sqrt{n}}\right) \\ &\le C'\varepsilon + \frac{C''}{\sqrt{n}}.\end{aligned}$$

For $\varepsilon > c_0$, the same conclusion follows trivially after increasing C'. All constants depend only on a, b, K, as claimed.

5 Gaussian Random Processes

Recall that a random vector is a collection of n random variables, forming a vector (i.e. $X = (X_1, \ldots, X_n)$). A discrete-time random process is a sequence of random variables; a random walk is an important example, obtained by adding successive random increments. More generally, a random process is a collection of random variables indexed by "time". Examples include

$$\{X_t \ : \ t \in \mathbb{R}\}, \quad \{X_t \ : \ t \ge 0\}, \quad \{X_t \ : \ t \in [a, b]\}.$$

5.1 Basic Concepts and Examples

Example 5.1 (Brownian Motion)**.** *A standard Brownian motion $\{X_t \ : \ t \ge 0\}$ is a random process with $X_0 = 0$ almost surely and the following properties:*

1. *For all $s \ge t \ge 0$ the random variable $X_s - X_t$ has distribution $N(0, s - t)$. This is known as having "Gaussian increments".*
2. *The function $f(t) = X_t$ is continuous in t almost surely.*
3. *Increments over disjoint time intervals are jointly independent. In particular, for $0 \le \gamma \le \tau \le t \le s$, the increments $X_s - X_t$ and $X_\tau - X_\gamma$ are independent.*

We will be interested in high dimensional random processes (i.e. random processes where "time" is a subset of $\mathbb{R}^n$).

Example 5.2 (Ocean temperature)**.** *To represent the temperature of the ocean as a random process we can take $T \subset \mathbb{R}^n$ and for $t \in T$ let X_t denote the temperature at t.*

Example 5.3 (Random Projection)**.** *Let $g \sim N(0, I_n)$ be a standard Gaussian vector. In other words, $g = (g_1, \ldots, g_n)$, where the coordinates g_i are independent standard normal random variables. Then for $t \in T \subset \mathbb{R}^n$ we define $X_t := \langle g, t \rangle$.*

We now recall the definition of a Gaussian random vector. Given a symmetric non-negative definite $n \times n$ matrix A and a vector $b \in \mathbb{R}^n$, the distribution $N(b, A)$ is the law of $b + A^{1/2}g$, where $g \sim N(0, I_n)$. If A is positive definite, this law has density

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det A}} \exp\left(-\frac{1}{2}\langle A^{-1}(x-b), x-b\rangle\right).$$

If A is singular, the distribution is supported on a proper affine subspace and does not have a density with respect to n-dimensional Lebesgue measure.

Remark 5.4. ***(Homework)*** *Let $X \sim N(0, \Sigma)$ be a Gaussian random vector where $\Sigma = AA^\top$. Then X has the same distribution as Ag, where $g \sim N(0, I_n)$. In particular, we may realize $X_i = \langle g, u_i\rangle$, where u_i is the ith row of A. Note that the coordinates of X may be dependent.*

Definition 5.5. The covariance matrix Σ of a random vector $X \in \mathbb{R}^n$ with finite second moments is the $n \times n$ matrix with entries

$$\Sigma_{ij} = \mathbb{E}[(X_i - \mathbb{E}\, X_i)(X_j - \mathbb{E}\, X_j)].$$

Remark 5.6. *If X has independent entries then Σ is a diagonal matrix.*

As a cool fact (HW), we remark that the distribution of a centered Gaussian vector $X \sim N(0, \Sigma)$ is uniquely determined by its covariance matrix. A general Gaussian law is determined by its mean and covariance. For non-Gaussian random vectors, even the mean and covariance usually do not determine the distribution; for example, a standard Gaussian variable and a symmetric random sign both have mean zero and variance one.

Definition 5.7 (Covariance Function)**.** Let $\{X_t : t \in T\}$ be a "mean zero" random process with finite second moments (i.e. $\mathbb{E}[X_t] = 0$ for all $t \in T$). We define $\Sigma\colon T \times T \to \mathbb{R}$ according to

$$\Sigma(t, s) = \mathbb{E}[X_t X_s].$$

Σ is known as the covariance function and is the random process analogue to the covariance matrix.

Definition 5.8 (Gaussian Random Process)**.** Let T be an index set. A random process $\{X_t : t \in T\}$ is called a Gaussian random process (GRP) if for every finite subset $T_0 \subset T$, the vector $(X_t)_{t\in T_0}$ is a Gaussian vector. An equivalent characterization is that for every finite subset $T_0 \subset T$ and real coefficients $(a_t)_{t\in T_0}$, the linear combination $\sum_{t\in T_0} a_t X_t$ is a Gaussian random variable (possibly with variance zero). This equivalence is because the projection of a Gaussian vector in any direction is a normal random variable, and a random vector whose projection in every direction is normal must be Gaussian. (HW) The finite-dimensional distributions of a centered Gaussian random process are determined by its covariance function.

Definition 5.9 ("Canonical GRP"). Let $T \subset \mathbb{R}^n$ and let $g \sim N(0, I_n)$. For every $t \in T$ define $X_t = \langle g, t \rangle$. Then $\{X_t \ : \ t \in T\}$ is known as a canonical Gaussian process. This is indeed a Gaussian random process: every finite linear combination has the form $\sum_t a_t X_t = \langle g, \sum_t a_t t \rangle$ and is therefore normal.

Lemma 5.10 (Finite Gaussian processes are canonical). *Let $\{Y_t : t \in T\}$ be a mean zero Gaussian random process with $|T| = m < \infty$. Then there exist vectors $u_t \in \mathbb{R}^m$ such that $(Y_t)_{t \in T}$ has the same distribution as $(\langle g, u_t \rangle)_{t \in T}$, where $g \sim N(0, I_m)$.*

Proof. Factor the covariance matrix as $AA^\top$ and take u_t to be the corresponding row of A, as in the homework above. □

For an infinite index set, the analogous representation uses a Gaussian process indexed by a Hilbert space; one cannot in general use a Gaussian vector in a fixed finite-dimensional space. The finite version above is sufficient for the comparison arguments below.

Remark 5.11 (Regularity conventions). *For the sake of presentation, we first argue with finite index sets and smooth functions for which the displayed expectations and differentiations are justified. The comparison results extend to processes whose suprema are almost surely determined by countable subsets (for example, continuous processes on separable index spaces), by taking increasing finite subsets. Smooth approximation and truncation handle the integration-by-parts steps whenever the required integrability holds. These conventions also apply to the sub-Gaussian processes below.*

5.2 Slepian's Inequality

In applications, it is useful to have a uniform control on a random process $\{X_t : t \in T\}$, i.e. to have a bound on $\mathbb{E} \sup_{t \in T} X_t$.

For some processes, this quantity can be computed exactly. For example, if $\{X_t : t \in T\}$ is a standard Brownian motion, then by the reflection principle, we have $\mathbb{E} \sup_{0 \le t \le t_0} X_t = \sqrt{2t_0/\pi}$ for every $t_0 \ge 0$. For general random processes, even if they are Gaussian, the problem is emphatically nontrivial.

The first general bound we will prove is Slepian's comparison inequality for Gaussian processes. Intuitively, it states that the faster the process grows (in terms of the magnitude of the increments), the farther it gets.

Theorem 5.12 (Slepian's Inequality). *Let $\{X_t \ : t \in T\}, \{Y_t \ : \ t \in T\}$ be mean zero Gaussian processes indexed by T, under the regularity convention above. Suppose that for all $s, t \in T$ it holds that*

$$\mathbb{E} X_t^2 = \mathbb{E} Y_t^2, \quad \mathbb{E}(X_t - X_s)^2 \le \mathbb{E}(Y_t - Y_s)^2.$$

Then for all $\tau \in \mathbb{R}$ it follows that

$$\mathbb{P}\left[\sup_{t \in T} X_t \ge \tau\right] \le \mathbb{P}\left[\sup_{t \in T} Y_t \ge \tau\right], \tag{47}$$

and thus

$$\mathbb{E}\sup_{t\in T} X_t \le \mathbb{E}\sup_{t\in T} Y_t. \tag{48}$$

Remark 5.13 (Homework)**.** *Instead of the Gaussian processes $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$, it suffices to prove the above inequality for the Gaussian random vectors X and Y in $\mathbb{R}^n$. Furthermore, it suffices to prove for the case when X and Y are independent. Hence, Theorem 5.12 is equivalent to Theorem 5.20 which will be proved later.*

For finite index sets, the inequalities (47) and (48) become

$$\mathbb{P}\left[\sup_{i\le n} X_i \ge \tau\right] \le \mathbb{P}\left[\sup_{i\le n} Y_i \ge \tau\right], \qquad \mathbb{E}\sup_{i\le n} X_i \le \mathbb{E}\sup_{i\le n} Y_i,$$

where the first inequality guarantees the second. Indeed, writing $M_X = \max_{i\le n} X_i$, the tail formula for an integrable real random variable gives

$$\mathbb{E}\, M_X = \int_0^\infty \mathbb{P}(M_X > r)\, dr - \int_0^\infty \mathbb{P}(M_X \le -r)\, dr.$$

The upper-tail comparison makes the first integral no larger for M_X than for M_Y, and the second integral no smaller. Thus $\mathbb{E}\, M_X \le \mathbb{E}\, M_Y$. Notice that the maxima need not be non-negative, so the formula with only one tail integral would not suffice.

5.2.1 Gaussian Interpolation

The proof of Slepian's inequality will be based on the technique of Gaussian Interpolation which is described as follows.

Definition 5.14 (Gaussian Interpolation)**.** For any pair of independent Gaussian random vectors $X, Y \in \mathbb{R}^n$, not necessarily standard, define a Gaussian random vector $Z(u)$ in $\mathbb{R}^n$ that continuously interpolates between $Z(0) = Y$ and $Z(1) = X$:

$$Z(u) := \sqrt{u}\, X + \sqrt{1-u}\, Y, \quad u \in [0,1].$$

Remark 5.15 (Homework)**.** *The covariance matrix of $Z(u)$ interpolates linearly between the covariance matrices of Y and X. Namely, if $\Sigma(X)$ is the covariance matrix for X and $\Sigma(Y)$ is the covariance matrix for Y, then*

$$\Sigma(Z(u)) = u\Sigma(X) + (1-u)\Sigma(Y)$$

Consider the indicator function for vector $x = (x_1, \cdots, x_n)$

$$f(x) := \mathbf{1}_{\{\max_i x_i < \tau\}}$$

which satisfies $\mathbb{E}\, f(Z(1)) = \mathbb{P}(\max_{i\le n} X_i < \tau)$ and $\mathbb{E}\, f(Z(0)) = \mathbb{P}(\max_{i\le n} Y_i < \tau)$. Now, if we can show

$$\mathbb{E}\, f(Z(1)) \ge \mathbb{E}\, f(Z(0))$$

then inequality (47) can be concluded. Our goal now shifts to study how the quantity $\mathbb{E}\, f(Z(u))$ changes as u increases from 0 to 1. We approach this goal by starting with the following identity which is a version of integration by parts in Gaussian expectations.

Lemma 5.16 (Gaussian integration by parts)**.** *Let $X \sim N(0,1)$. Then for any continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ satisfying the integrability conditions discussed above, we have*

$$\mathbb{E}\, f'(X) = \mathbb{E}\, X f(X) \tag{49}$$

Proof. It suffices to argue for the case when f has compact support; the identity extends by smooth cutoffs when $\mathbb{E}\,|f'(X)|$ and $\mathbb{E}\,|Xf(X)|$ are finite. By the density of the standard normal distribution and integration by parts, we have

$$\mathbb{E}\, f'(X) = \int_{\mathbb{R}} f'(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 0 + \int_{\mathbb{R}} t f(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \mathbb{E}\, X f(X)$$

as claimed, where the zero term comes from the fact that $f(t)e^{-t^2/2}$ has limits equal to zero when t goes to both positive and negative infinity if f has bounded support. □

Corollary 5.17 (Homework: Multivariate Gaussian Integration by parts)**.** *Let $X \sim N(0, \Sigma)$. Then, under the same regularity convention, for a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathbb{E}\, X f(X) = \Sigma\ \ \mathbb{E}\, \nabla f(X),$$

where $\nabla f(X)$ is the n-dimensional gradient vector with entries $\partial_i f(X)$.

Lemma 5.18 (Gaussian Interpolation)**.** *Consider two independent Gaussian random vectors $X \sim N(0, \Sigma^X)$ and $Y \sim N(0, \Sigma^Y)$. Define the interpolation Gaussian vector*

$$Z(u) := \sqrt{u} X + \sqrt{1-u} Y, \qquad u \in [0,1]. \tag{50}$$

Then for a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying the regularity convention above, and $0 < u < 1$, we have

$$\frac{d}{du}\, \mathbb{E}\, f(Z(u)) = \frac{1}{2} \sum_{i,j=1}^{n} \left(\Sigma^X_{ij} - \Sigma^Y_{ij} \right) \mathbb{E}\, \frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)), \tag{51}$$

where Σ^X_{ij} stands for the (i,j) entry of Σ^X.

Proof. By the (multivariate) chain rule, we have

$$\frac{d}{du}\, \mathbb{E}\, f(Z(u)) = \sum_{i=1}^{n} \mathbb{E}\, \frac{\partial f}{\partial x_i}(Z(u)) \frac{dZ_i}{du} = \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\, \frac{\partial f}{\partial x_i}(Z(u)) \left(\frac{X_i}{\sqrt{u}} - \frac{Y_i}{\sqrt{1-u}} \right),$$

where the second equality is from the definition (50) of $Z(u)$. Now break this sum into two, and first compute the contribution of terms containing X_i. To this end, we condition on Y and express

$$\sum_{i=1}^{n} \frac{1}{\sqrt{u}}\, \mathbb{E}\, X_i \frac{\partial f}{\partial x_i}(Z(u)) = \sum_{i=1}^{n} \frac{1}{\sqrt{u}}\, \mathbb{E}\, X_i g_i(X),$$

where

$$g_i(X) = \frac{\partial f}{\partial x_i}\left(\sqrt{u}X + \sqrt{1-u}Y\right).$$

Applying multivariate Gaussian integration by parts (Corollary 5.17) conditionally on Y, and then averaging over Y, we have

$$\mathbb{E}\, X_i g_i(X) = \sum_{j=1}^n \Sigma_{ij}^X\, \mathbb{E}\, \frac{\partial g_i}{\partial x_j}(X) = \sqrt{u}\sum_{j=1}^n \Sigma_{ij}^X\, \mathbb{E}\, \frac{\partial^2 f}{\partial x_i \partial x_j}\left(\sqrt{u}X + \sqrt{1-u}Y\right),$$

where the second equality is by definition of g_i. Substituting this back into the previous equation gives

$$\sum_{i=1}^n \frac{1}{\sqrt{u}}\, \mathbb{E}\, X_i \frac{\partial f}{\partial x_i}(Z(u)) = \sum_{i,j=1}^n \Sigma_{ij}^X\, \mathbb{E}\, \frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)).$$

Similarly, conditioning on X handles the contribution of terms containing Y_i and yields

$$\sum_{i=1}^n \frac{1}{\sqrt{1-u}}\, \mathbb{E}\, Y_i \frac{\partial f}{\partial x_i}(Z(u)) = \sum_{i,j=1}^n \Sigma_{ij}^Y\, \mathbb{E}\, \frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)).$$

Combining these two equalities, together with the first line of the proof, gives the desired relation (51). □

5.2.2 Proof of Slepian's Inequality

We are ready to establish the key lemma in proving Slepian's inequality which is also known as a preliminary functional form of Slepian's inequality.

Lemma 5.19 (Slepian's inequality: functional form)**.** *Consider two mean-zero independent Gaussian random vectors X and Y in $\mathbb{R}^n$. Assume that for all $i, j = 1, \cdots, n$, we have*

$$\mathbb{E}\, X_i^2 = \mathbb{E}\, Y_i^2 \qquad \text{and} \qquad \mathbb{E}(X_i - X_j)^2 \le \mathbb{E}(Y_i - Y_j)^2. \tag{52}$$

Consider a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, under the regularity convention above, such that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \ge 0 \quad \text{for all } i \ne j.$$

Then

$$\mathbb{E}\, f(X) \ge \mathbb{E}\, f(Y).$$

Proof. The assumptions (52) imply that the entries of the covariance matrices Σ^X and Σ^Y of X and Y satisfy

$$\Sigma_{ii}^X = \Sigma_{ii}^Y \qquad \text{and} \qquad \Sigma_{ij}^X \ge \Sigma_{ij}^Y, \quad \text{for all } i, j = 1, \cdots, n,$$

where for the second relation we used $ab = (a^2 + b^2 - (a-b)^2)/2$. Applying Lemma 5.18 with our assumptions gives

$$\frac{d}{du}\,\mathbb{E}\, f(Z(u)) = \frac{1}{2}\sum_{i,j=1}^{n} \left(\Sigma_{ij}^X - \Sigma_{ij}^Y\right) \mathbb{E}\,\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)) \geq 0$$

by the assumptions, where $Z(u)$ is the Gaussian interpolation we defined in (50). This means $\mathbb{E}\, f(Z(u))$ increases in u. But $Z(0) = Y$ and $Z(1) = X$ by the way we constructed the interpolation, hence $\mathbb{E}\, f(X) = \mathbb{E}\, f(Z(1)) \geq \mathbb{E}\, f(Z(0)) = \mathbb{E}\, f(Y)$ as desired. □

We are now ready to prove Slepian's inequality (47) in its finite-dimensional form, in terms of random vectors rather than random processes.

Theorem 5.20 (Slepian's inequality: random vector)**.** *Let X and Y be two mean zero independent Gaussian random vectors in $\mathbb{R}^n$. Suppose that for all $i, j \in \{1, \ldots, n\}$ it holds that*

$$\mathbb{E}\, X_i^2 = \mathbb{E}\, Y_i^2, \quad \mathbb{E}(X_i - X_j)^2 \leq \mathbb{E}(Y_i - Y_j)^2.$$

Then for every $\tau \in \mathbb{R}$ we have

$$\mathbb{P}\left[\max_{i\leq n} X_i \geq \tau\right] \leq \mathbb{P}\left[\max_{i\leq n} Y_i \geq \tau\right]. \tag{53}$$

Consequently,

$$\mathbb{E}\max_{i\leq n} X_i \leq \mathbb{E}\max_{i\leq n} Y_i. \tag{54}$$

Proof. Let $h_\epsilon : \mathbb{R} \to [0,1]$ be a smooth non-increasing function equal to 1 on $(-\infty, \tau - \epsilon]$ and 0 on $[\tau, \infty)$. Thus h_ϵ converges pointwise to $\mathbf{1}_{(-\infty,\tau)}$ as $\epsilon \downarrow 0$. Define $f_\epsilon(x) = h_\epsilon(x_1) \cdots h_\epsilon(x_n)$. Then

$$f_\epsilon(x) \longrightarrow \mathbf{1}_{\{\max_i x_i < \tau\}}.$$

To apply the functional form of Slepian's inequality, we need to check the assumption for f_ϵ. Note that, for every $i \neq j$, we have

$$\frac{\partial^2 f_\epsilon}{\partial x_i \partial x_j} = h_\epsilon'(x_i) h_\epsilon'(x_j) \prod_{k \notin \{i,j\}} h_\epsilon(x_k).$$

The first two factors are non-positive and the others are non-negative. Thus every mixed second derivative with $i \neq j$ is non-negative. The functional form of Slepian's inequality gives

$$\mathbb{E}\, f_\epsilon(X) \geq \mathbb{E}\, f_\epsilon(Y).$$

Letting $\epsilon \downarrow 0$ and using bounded convergence, we obtain

$$\mathbb{P}\left(\max_{i\leq n} X_i < \tau\right) = \lim_{\epsilon\downarrow 0} \mathbb{E}\, f_\epsilon(X) \geq \lim_{\epsilon\downarrow 0} \mathbb{E}\, f_\epsilon(Y) = \mathbb{P}\left(\max_{i\leq n} Y_i < \tau\right)$$

which implies (53). And relation (54) follows from (53) as we discussed before. □

5.2.3 The Sudakov-Fernique Inequality

In Theorem 5.12, Slepian's inequality has two assumptions on the random processes $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$: equality of variances and dominance of increments. It turns out that, even if we drop equality of variances, we can still obtain the inequality on expectations. This more practically useful result is due to Sudakov and Fernique.

Theorem 5.21 (Sudakov-Fernique Inequality)**.** *Let $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be two mean zero Gaussian processes, under the regularity convention above. Assume that, for all $t, s \in T$, we have*

$$\mathbb{E}(X_t - X_s)^2 \leq \mathbb{E}(Y_t - Y_s)^2.$$

Then

$$\mathbb{E} \sup_{t \in T} X_t \leq \mathbb{E} \sup_{t \in T} Y_t.$$

Proof. It is enough to prove this theorem for Gaussian random vectors X and Y in $\mathbb{R}^n$, which we may take to be independent, just as for Slepian's inequality. We again deduce the result from the Gaussian interpolation lemma 5.18. This time, instead of choosing a function that approximates the indicator of $\{\max_i x_i < \tau\}$, we want to approximate $\max_i x_i$.

To this end, let $\beta > 0$ be a parameter and define the function

$$f_\beta(x) := \frac{1}{\beta} \log \sum_{i=1}^{n} e^{\beta x_i}.$$

One can check (Homework!) that this function is smooth and

$$\max_{i \leq n} x_i \leq f_\beta(x) \leq \max_{i \leq n} x_i + \frac{\log n}{\beta}.$$

Put $p_i(x) = e^{\beta x_i} / \sum_k e^{\beta x_k}$. Then $\partial_i f_\beta = p_i$ and $\partial_{ij} f_\beta = \beta(p_i \delta_{ij} - p_i p_j)$. Substituting into the Gaussian interpolation formula and using $\sum_i p_i = 1$ gives

$$\frac{d}{du} \mathbb{E}\, f_\beta(Z(u)) = \frac{\beta}{4} \sum_{i,j} \mathbb{E}\left[p_i(Z(u)) p_j(Z(u))\right] \left(\mathbb{E}(X_i - X_j)^2 - \mathbb{E}(Y_i - Y_j)^2\right) \leq 0.$$

Hence $\mathbb{E}\, f_\beta(X) \leq \mathbb{E}\, f_\beta(Y)$. Letting $\beta \to \infty$ in the uniform approximation above proves the desired inequality. □

5.2.4 Application of Sudakov-Fernique Inequality to Gaussian Random Matrices

Theorem 5.22. *Let A be an $m \times n$ matrix with entries (a_{ij}) such that a_{ij} are independent and $a_{ij} \sim \mathcal{N}(0, 1)$. Then we have*

$$\mathbb{E}\|A\| \leq \sqrt{m} + \sqrt{n}.$$

Remark 5.23. *Note that $\|A\|_{op} = \sigma_1(A)$. This is consistent with our earlier bounds of order $\sqrt{m} + \sqrt{n}$ for matrices with independent, centered, uniformly sub-Gaussian entries. For standard Gaussian entries, the comparison argument gives the precise leading constant 1 in this expectation bound.*

Proof of Theorem 5.22. Let A be an $m \times n$ matrix with entries (a_{ij}) such that a_{ij} are independent and $a_{ij} \sim \mathcal{N}(0,1)$. Then

$$\|A\| = \sup_{x \in \mathbb{S}^{n-1}} |Ax| = \sup_{\substack{y \in \mathbb{S}^{m-1} \\ x \in \mathbb{S}^{n-1}}} \langle Ax, y \rangle = \max_{\substack{t=(x,y) \\ t \in T \\ T = \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}}} X_t,$$

since $|a| = \sup_{y \in \mathbb{S}^{m-1}} \langle a, y \rangle$ for $a \in \mathbb{R}^m$. Here $X_{(x,y)} = \langle Ax, y \rangle$ is a centered Gaussian process: every finite linear combination of its values is a linear combination of the independent Gaussian entries of A. The idea is to find a simpler Gaussian process Y_t on the same set $T = \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$ whose increments dominate those of X_t. The Sudakov-Fernique inequality will then give $\mathbb{E}\|A\| = \mathbb{E} \sup X_t \leq \mathbb{E} \sup Y_t$.

Let $t = (u, v) \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$ and let $s = (w, z) \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$. Then

$$\begin{aligned}
\mathbb{E}\left(X_t - X_s\right)^2 &= \mathbb{E}\left(\langle Au, v \rangle - \langle Aw, z \rangle\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\left(u_j v_i - w_j z_i\right)\right)^2 \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{E} a_{ij}^2 \left(u_j v_i - w_j z_i\right)^2 \\
&= \|v \otimes u - z \otimes w\|_{HS}^2 \\
&= 2 - 2\langle u, w \rangle \langle v, z \rangle \\
&\leq |u - w|^2 + |v - z|^2.
\end{aligned}$$

Here $(v \otimes u)_{ij} = v_i u_j$. The last inequality follows because the right-hand side minus the preceding expression is $2(1 - \langle u, w \rangle)(1 - \langle v, z \rangle) \geq 0$. Recall also that for independent, mean zero $\xi_1, \ldots, \xi_k$, we have $\mathbb{E}(\sum_l \xi_l)^2 = \sum_l \mathbb{E}\xi_l^2$; this applies to the summands $a_{ij}(u_j v_i - w_j z_i)$. We conclude that

$$\mathbb{E}|X_{uv} - X_{wz}|^2 \leq |u - w|^2 + |v - z|^2.$$

In fact, we can construct Y_t with $t \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$ such that

$$\mathbb{E}|Y_{uv} - Y_{wz}|^2 = |u - w|^2 + |v - z|^2.$$

Indeed, consider $Y_{uv} = \langle g, u \rangle + \langle h, v \rangle$ where $h \sim \mathcal{N}(0, I_m)$, $g \sim \mathcal{N}(0, I_n)$, h, g are independent, and $(u, v) \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$. Note that here Y_{uv} is a Gaussian random process, then by definition

$$\begin{aligned}
\mathbb{E}|Y_{uv} - Y_{wz}|^2 &= \mathbb{E}\left|\langle g, u \rangle + \langle h, v \rangle - \langle g, w \rangle - \langle h, z \rangle\right|^2 \\
&= \mathbb{E}\langle g, u - w \rangle^2 + \mathbb{E}\langle h, v - z \rangle^2 \qquad \text{(by independence and mean zero)} \\
&= |u - w|^2 + |v - z|^2
\end{aligned}$$

since if $G \sim \mathcal{N}(0, I_d)$, then $\langle G, \theta \rangle \sim \mathcal{N}(0, |\theta|^2)$ for $\theta \in \mathbb{R}^d$. By the results above, $\mathbb{E}|X_t - X_s|^2 \leq \mathbb{E}|Y_t - Y_s|^2$. Since both are centered Gaussian processes, the Sudakov-Fernique inequality gives

$$\begin{aligned}
\mathbb{E}\|A\|_{op} = \mathbb{E} \sup_{t \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}} X_t &\leq \mathbb{E} \sup_{t \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}} Y_t \\
&= \mathbb{E} \sup_{\substack{u \in \mathbb{S}^{n-1} \\ v \in \mathbb{S}^{m-1}}} (\langle g, u \rangle + \langle h, v \rangle) \\
&= \mathbb{E} \sup_{u \in \mathbb{S}^{n-1}} \langle g, u \rangle + \mathbb{E} \sup_{v \in \mathbb{S}^{m-1}} \langle h, v \rangle \\
&= \mathbb{E}|g| + \mathbb{E}|h| \\
&\overset{\text{Cauchy}}{\leq} \sqrt{\mathbb{E}|g|^2} + \sqrt{\mathbb{E}|h|^2} \\
&= \sqrt{n} + \sqrt{m}
\end{aligned}$$

as $\mathbb{E}|g|^2 = \mathbb{E}\left(\sum_{i=1}^{n} g_i^2\right) = n$ and $g \sim \mathcal{N}(0, I_n)$. □

Definition 5.24 (Sub-Gaussian Random Process)**.** For a process $\{X_t : t \in T\}$ and a metric d on T, we say that the process has *sub-Gaussian increments* (or is sub-Gaussian with respect to d) if, for some constant $K \geq 0$,

$$\|X_t - X_s\|_{\psi_2} \leq K \cdot d(t, s),$$

i.e., all increments are sub-Gaussian random variables, with their norms controlled by the distance between their indices.

Every centered Gaussian process has sub-Gaussian increments with respect to its canonical distance $d(s,t) = (\mathbb{E}\,|X_t - X_s|^2)^{1/2}$, with an absolute constant K. This distance may be a pseudometric; identifying points at distance zero makes it a metric.

Theorem 5.25 (Dudley's Inequality)**.** *Suppose $\{X_t : t \in T\}$ is a mean zero random process on a metric space (T, d) satisfying the regularity convention above, and has sub-Gaussian increments with constant K. Then*

$$\mathbb{E} \sup_{t \in T} X_t \leq CK \sum_{j \in \mathbb{Z}} 2^{-j} \sqrt{\log N\,(T, d, 2^{-j})},$$

where $\log N\,(T, d, 2^{-j})$ is the metric entropy, i.e., $N\,(T, d, 2^{-j})$ is the smallest number of balls of radius 2^{-j} required to cover T in metric d.

Remark 5.26. *If T is nonempty and bounded (in particular, if it is compact), there is $j_0 \in \mathbb{Z}$ such that $N(T, d, 2^{-j}) = 1$ for all $j \leq j_0$, so all corresponding summands are zero.*

Example 5.27. *Let $T = \mathbb{S}^{n-1} \subset \mathbb{R}^n$ and let d be the Euclidean metric. The volumetric covering bound gives $N(\mathbb{S}^{n-1}, d, \epsilon) \leq (1 + 2/\epsilon)^n$, and hence $N(\mathbb{S}^{n-1}, d, \epsilon) \leq (3/\epsilon)^n$ for $0 < \epsilon \leq 1$. Also, $N(\mathbb{S}^{n-1}, d, \epsilon) = 1$ for $\epsilon \geq 2$. Therefore*

$$\begin{aligned} C \sum_{j \in \mathbb{Z}} 2^{-j} \sqrt{\log N\left(\mathbb{S}^{n-1}, d, 2^{-j}\right)} &\leq C\sqrt{n} \sum_{j \geq 0} 2^{-j} \sqrt{\log 3 + j \log 2} \\ &\leq C'\sqrt{n} \sum_{j \geq 0} 2^{-j} \sqrt{j+1} \leq C''\sqrt{n}. \end{aligned}$$

In conclusion, $\mathbb{E} \sup_{t \in \mathbb{S}^{n-1}} X_t \leq C\sqrt{n}$ for a mean zero process with 1-sub-Gaussian increments in the Euclidean metric. Dudley's inequality also applies to processes indexed by the Hamming cube, once an increment bound in the chosen metric has been verified.

6 The Semigroup method

The semigroup method is a powerful method for proving "isoperimetric-type" inequalities.

Remark 6.1 (Regularity conventions)**.** *For the sake of presentation, we first work with smooth functions for which differentiation under the integral and integration by parts are justified. The resulting inequalities extend to their natural domains by truncation, smoothing, and passage to the limit. In abstract semigroup arguments we assume strong continuity on $L^2(\mu)$ and work on the generator domain when differentiating. Except in the explicitly mentioned Lebesgue-measure example, stationary measures below are probability measures.*

6.1 Basic definitions and set up

Definition 6.2 (Markov Process)**.** Consider a real-valued random process $(X_t)_{t \in T}$, where $T \subset \mathbb{R}$, and times $t_1 < \cdots < t_n$. A *Markov process* is a stochastic process with the property that

$$\mathbb{P}\left(X_{t_n} \leq x_n \,|\, X_{t_{n-1}}, X_{t_{n-2}}, \ldots, X_{t_1}\right) = \mathbb{P}\left(X_{t_n} \leq x_n \,|\, X_{t_{n-1}}\right),$$

where $x_n \in \mathbb{R}$ and $X_t \in \mathbb{R}$, i.e., this stochastic process "does not see the past".

Recall that, when $\mathbb{P}(B) > 0$, the conditional probability is $\mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, and the conditional expectation $\mathbb{E}(X \,|\, Y)$ of random variables X and Y is also a random variable which is the best prediction of X given some behavior of Y. Suppose (X, Y) has a joint density f_{XY}, with marginal densities f_X and f_Y. At points where $f_Y(y) > 0$, we can define the conditional density as $f_{X \,|\, Y}(x \,|\, y) = \frac{f_{XY}(x,y)}{f_Y(y)}$, so $\mathbb{E}\left(X \,|\, Y = y\right) = \int_{-\infty}^{\infty} x \cdot f_{X \,|\, Y}(x \,|\, y)\, dx$ which is a number that depends on y. Substituting the random variable Y for y gives the random variable $\mathbb{E}(X \,|\, Y)$.

Definition 6.3 (Conditional Expectation)**.** Let X be an integrable random variable on $(\Omega, \mathcal{F}_0, \mathbb{P})$, and let $\mathcal{F} \subseteq \mathcal{F}_0$ be a sub-σ-algebra. The *conditional expectation of X with respect*

to $\mathcal{F}$, denoted by $\mathbb{E}(X \mid \mathcal{F})$, is an integrable, $\mathcal{F}$-measurable random variable such that, for every $A \in \mathcal{F}$,

$$\int_A X \, d\mathbb{P} = \int_A \mathbb{E}(X \mid \mathcal{F}) \, d\mathbb{P}.$$

It is unique almost surely. For an integrable random vector, apply this definition coordinate-wise. If Y is another random vector, then the conditional expectation of X with respect to Y is

$$\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \sigma(Y)),$$

where $\sigma(Y)$ is the σ-algebra generated by Y.

Definition 6.4 (Time-homogeneous Markov Process)**.** Consider $X_t \in \mathbb{R}^n$ with $t \geq 0$ as a random process. Assume for any bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$ and for any times $t, s \geq 0$, there is a bounded measurable function $P_s f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbb{E}(f(X_{t+s}) \mid \{X_z\}_{z \leq t}) = P_s f(X_t), \tag{55}$$

i.e., the conditional law of X_{t+s} only depends on X_t. The notation P_s also assumes time homogeneity: the transition rule depends on the elapsed time s, not on the starting time t. We use this assumption throughout.

Homework: find an example of a non-Markov process.

Definition 6.5 (Markov Semigroup)**.** Suppose P_s is an operator on bounded measurable functions such that

$$f : \mathbb{R}^n \to \mathbb{R} \quad \xrightarrow{P_s} \quad P_s f : \mathbb{R}^n \to \mathbb{R}$$

as defined in (55). The family $(P_s)_{s \geq 0}$ is called a *Markov semigroup*; P_0 is the identity. Each P_s preserves nonnegativity and constants, and the semigroup property is proved below.

Definition 6.6 (Stationary measure of a Markov Process)**.** Consider X_t on $\mathbb{R}^n$ as a Markov process indexed by $T \subset \mathbb{R}^+$. A probability measure μ on $\mathbb{R}^n$ is called a *stationary measure of* X_t if for any bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$\int f \, d\mu = \int P_t f \, d\mu.$$

For fixed f, the functions $P_t f$ describe its evolution in time. The measure is stationary if this evolution preserves the integral of every f.

Remark 6.7. *Consider initial random vector $X_0 \sim \mu$ where μ is a stationary measure. Then*

$$\begin{aligned}
\mathbb{E}f(X_t) &= \mathbb{E}(\mathbb{E}(f(X_t) \mid X_0)) \\
&= \mathbb{E}P_t f(X_0) \\
&= \int P_t f \, d\mu \qquad (\textit{as } X_0 \sim \mu) \\
&= \int f \, d\mu \qquad (\textit{by stationary}) \\
&= \mathbb{E}f(X_0).
\end{aligned}$$

Hence, for every bounded measurable function f, we have $\mathbb{E}f(X_t) = \mathbb{E}f(X_0)$.
If $f = \mathbb{1}_\Omega$, then $\mathbb{P}(X_t \in \Omega) = \mathbb{P}(X_0 \in \Omega)$. In other words, if $X_0 \sim \mu$ with stationary measure μ, then for any $t \geq 0$, $X_t \sim \mu$.

6.2 Properties of Markov semigroups, and some examples

Lemma 6.8. *Let μ be a stationary measure of a Markov process X_t indexed by $t \geq 0$. Then the following hold for all $1 \leq p \leq \infty$, $\alpha, \beta \in \mathbb{R}$, bounded measurable function f, g:*

1. *$\|P_t f\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)}$ (contraction).*
2. *P_t is a linear operator, i.e. $P_t(\alpha f + \beta g) = \alpha P_t f + \beta P_t g$ (linearity).*
3. *$P_{t+s} f = P_t P_s f \quad \mu$-a.s. (semigroup property).*
4. *$P_t 1 = 1 \quad \mu$-a.s. (conservativeness).*

Proof. 1. Suppose $X_0 \sim \mu$, we have

$$\int |P_t f|^p \, d\mu = \mathbb{E}\left|\, \mathbb{E}(f(X_t) \mid X_0)\right|^p \leq \mathbb{E}\,|f(X_t)|^p = \int |f|^p \, d\mu$$

for $p < \infty$, by conditional Jensen's inequality and stationarity. For $p = \infty$, use $|P_t f| \leq P_t |f| \leq \|f\|_\infty$.

2. Linearity follows similarly as

$$\begin{aligned} P_t(\alpha f + \beta g)(X_0) &= \mathbb{E}(\alpha f(X_t) + \beta g(X_t) \mid X_0) \\ &= \alpha\, \mathbb{E}(f(X_t) \mid X_0) + \beta\, \mathbb{E}(g(X_t) \mid X_0) \\ &= \alpha P_t f(X_0) + \beta P_t g(X_0), \end{aligned}$$

by linearity of conditional expectation.

3. For the semigroup property, we have

$$\begin{aligned} P_{t+s} f(X_0) &= \mathbb{E}(f(X_{t+s}) \mid X_0) \\ &= \mathbb{E}\left(\mathbb{E}(f(X_{t+s}) \mid \{X_r\}_{r \leq t}) \mid X_0\right) \\ &= \mathbb{E}(P_s f(X_t) \mid X_0) = P_t P_s f(X_0), \end{aligned}$$

where the third equality is based on the Markov property.

4. Conservativeness follows $P_s 1 = \mathbb{E}(1 | X_0) = 1$.

□

Remark 6.9. *The contraction operators extend by continuity to every $L^p(\mu)$, $1 \leq p < \infty$. Linearity, preservation of constants, and the semigroup property continue to hold. We use these extensions when f is unbounded.*

Remark 6.10. *Not every Markov process has a stationary measure; the questions of existence are complicated. We will consider concrete examples when things work well, the stationary measure exists and has nice properties, and we will explain the existence in these examples.*

Definition 6.11 (Variance)**.** Let μ be a stationary measure and $f \in L^2(\mu)$. The variance is defined as

$$\mathrm{Var}_\mu(f) := \int f^2 d\mu - \left(\int f d\mu\right)^2 = \mathbb{E}_\mu f^2 - (\mathbb{E}_\mu f)^2. \tag{56}$$

Note, by Cauchy's inequality, that $\mathrm{Var}_\mu(f) \geq 0$.

Lemma 6.12. *Let μ be a stationary measure of a Markov process X_t and $f \in L^2(\mu)$. Then $\mathrm{Var}_\mu(P_t f)$ is nonincreasing in $t \geq 0$.*

Proof. Note that

$$\mathrm{Var}_\mu(P_t f) = \int (P_t f - \int P_t f d\mu)^2 d\mu = \int (P_t f - \int f d\mu)^2 d\mu = \int (P_t(f - \int f d\mu))^2 d\mu \tag{57}$$

$$\leq \int (f - \int f d\mu)^2 d\mu = \mathrm{Var}_\mu(f). \tag{58}$$

Here we use the definition of stationary measure for the second equality, linearity and conservativeness for the third equality and contraction for the fourth. Also we have

$$\int (P_t(f - \int f d\mu))^2 d\mu = \int (P_{t-s} P_s(f - \int f d\mu))^2 d\mu \leq \mathrm{Var}_\mu(P_s f) \tag{59}$$

Hence, we get $\mathrm{Var}_\mu(P_t f) \leq \mathrm{Var}_\mu(P_s f)$ for all $t \geq s \geq 0$. □

Remark 6.13. $\mathrm{Var}_\mu(f)$ *measures how far is f from a constant, as*

$$\mathrm{Var}_\mu(C) = \int C^2 d\mu - (\int C d\mu)^2 = C^2 - C^2 = 0.$$

Therefore, variance is a measure of the distance of the function to a constant function. The fact that it decreases along the semi-group means that the function $P_t f$ becomes closer and closer to a constant function – namely, to the function $\int f d\mu$ (since this integral is preserved). Soon we will see that in some nice situations, not only does the variance decrease, but it decreases all the way to zero, and $P_t f \underset{t\to\infty}{\to} \int f d\mu$; however, this is not necessarily the case for an arbitrary Markov process.

Definition 6.14 (Generator of Markov process)**.** Given a Markov process X_t with stationary measure μ on $\mathbb{R}^n$. The generator $\mathscr{L}$ is defined on those $f \in L^2(\mu)$ for which the following limit exists in $L^2(\mu)$:

$$\mathscr{L} f := \lim_{t \downarrow 0} \frac{P_t f - f}{t}. \tag{60}$$

Here, $\mathscr{L}$ is an operator on functions from $L^2(\mu)$ for which this limit makes sense. $\mathscr{L}$ is called the Generator associated with X_t.

Remark 6.15. *$\mathscr{L}$ is a linear operator, since $P_t f$ is linear.*

Remark 6.16 (Important)**.** *For f in the generator domain, the semigroup solves an evolution equation. Indeed, $\frac{d}{dt}P_t f = \lim_{\delta\to 0}\frac{P_{t+\delta}f - P_t f}{\delta} = \lim_{\delta\to 0} P_t(\frac{P_\delta f - f}{\delta}) = P_t\mathscr{L}f$. Equivalently, $\frac{d}{dt}P_t f = \lim_{\delta\to 0}\frac{P_\delta P_t f - P_t f}{\delta} = \mathscr{L}P_t f$.*

Corollary 6.17.

$$\mathscr{L}P_t f = P_t\mathscr{L}f. \tag{61}$$

Consider the following PDE:

$$\begin{cases} \dfrac{d}{dt}(P_t f) = \mathscr{L}(P_t f), \\ P_0 f = f, \end{cases}$$

Conversely, when this initial-value problem is well posed and its solution operators preserve nonnegativity and constants, it defines a Markov semigroup. An arbitrary linear operator need not have these properties. In our examples, PDE and probabilistic methods describe the same evolution.

Example 6.18 (Finite state space)**.** *Let $(X_t)_{t\in\mathbb{R}_+}$ be a Markov process with values in a finite state space $X_t \in \{1,\dots,d\}$. Such processes are typically described in terms of their transition rates $\lambda_{ij} \geq 0$ for $i \neq j$:*

$$\mathbb{P}[X_{t+\delta} = j \mid X_t = i] = \lambda_{ij}\delta + o(\delta) \quad \text{for } i \neq j.$$

Evidently, the transition rates λ_{ij} describe the infinitesimal rate of growth of the probability of jumping from state i to state j (informally if $X_t = i$ then the probability that $X_{t+dt} = j$ is $\lambda_{ij}dt$).

Let us organize the transition probabilities $q_{tij} = \mathbb{P}[X_t = j \mid X_0 = i]$ and rates λ_{ij} into matrices $Q_t = (q_{tij})_{1\leq i,j\leq d}$ and $\Lambda = (\lambda_{ij})_{1\leq i,j\leq d}$ respectively where we define the diagonal entries of Λ as $\lambda_{ii} = -\sum_{j\neq i}\lambda_{ij} \leq 0$. Then

$$\lim_{t\to 0}\frac{q_{tij} - q_{0ij}}{t} = \lambda_{ij}$$

for every $1 \leq i, j \leq d$ (the diagonal entries λ_{ii} were chosen precisely to enforce the law of total probability $\sum_j q_{tij} = 1$). In particular, we have

$$\mathscr{L}f(i) = \lim_{t\downarrow 0}\sum_{j=1}^{d}\frac{q_{tij} - q_{0ij}}{t}f(j) = \sum_{j=1}^{d}\lambda_{ij}f(j) = (\Lambda f)_i$$

where we identify the function f with the vector $(f(1),\dots,f(d)) \in \mathbb{R}^d$. We therefore conclude that the generator of a Markov process in a finite state space corresponds precisely to the matrix of transition rates. The Kolmogorov equation now reduces to the matrix differential equation (semigroup)

$$\frac{d}{dt}Q_t = Q_t\Lambda, \quad Q_0 = I.$$

This differential equation is the basic tool for computing probabilities of finite state space Markov processes. The solution is in fact easily obtained as $Q_t = e^{t\Lambda}$ from which we readily see why P_t and $\mathscr{L}$ must commute.

Example 6.19 (Heat semigroup)**.** *Let $\mathscr{L} = \Delta$ be the Laplace operator and $u : \mathbb{R}^n \to \mathbb{R}$.*

$$\Delta u = \sum_{i=1}^{n} \partial_{ii} u = \operatorname{tr}(\nabla^2 u). \tag{62}$$

Here, $\partial_{ii}u$ is the partial derivative and $\nabla^2 u$ is the Hessian matrix. The $L^2(dx)$ generator domain is $H^2(\mathbb{R}^n)$. For our calculations, smooth rapidly decaying functions suffice. The heat semigroup solves

$$\begin{cases} \partial_t(P_t f) = \Delta(P_t f), \\ P_0 f = f. \end{cases}$$

The solution is convolution with the heat kernel,

$$P_t f(x) = \int_{\mathbb{R}^n} f(x+y) \frac{e^{-|y|^2/(4t)}}{(4\pi t)^{n/2}}\, dy,$$

and corresponds to Brownian motion with generator Δ. What is the invariant measure? The condition is

$$\frac{d}{dt}\int P_t f \, d\mu = \int \Delta(P_t f)\, d\mu = 0.$$

For smooth functions with sufficient decay, Green's formula gives

$$\int u \Delta v \, dx = -\int \langle \nabla u, \nabla v \rangle \, dx.$$

Taking $u = 1$ gives $\int \Delta v \, dx = 0$. Thus Lebesgue measure is invariant for the heat semigroup, as also follows from the heat-kernel formula and Fubini's theorem. This invariant measure is infinite, rather than a stationary probability measure on $\mathbb{R}^n$.

Example 6.20 (Heat semigroup on the circle/torus)**.** *Consider a nice enough one dimensional function $f : [-\pi, \pi] \to \mathbb{R}$ (2π periodic). Then we have the PDE*

$$\begin{cases} \partial_t(P_t f) = \Delta(P_t f) \\ P_0 f = f \end{cases}$$

Then the stationary measure is uniform on circle.

Example 6.21 (Ornstein–Uhlenbeck semigroup)**.** *The Ornstein–Uhlenbeck operator is the second-order linear operator*

$$\mathscr{L}u = \Delta u - \langle \nabla u, x \rangle$$

Ornstein–Uhlenbeck semigroup is defined as: $\frac{d}{dt}(P_t f) = \mathscr{L}(P_t f)$, $P_0 f = f$. *The question here is: What is the stationary measure for this process? Observe that*

$$\int P_t f d\mu = \int f d\mu, \tag{63}$$

$$\frac{d}{dt} \int P_t f d\mu = 0, \tag{64}$$

$$\int \frac{d}{dt}(P_t f) d\mu = \int \mathscr{L}(P_t f) d\mu = 0. \tag{65}$$

The answer is the Gaussian measure!!!

$$d\gamma = \frac{1}{\sqrt{2\pi}^n} \exp -\frac{|x|^2}{2} dx, \tag{66}$$

$$\int u \mathscr{L} v d\gamma = -\int \langle \nabla u, \nabla v \rangle d\gamma, \tag{67}$$

for nice enough u, v. *Plugging in* $u = 1$ *gives* $\int \mathscr{L} v \, d\gamma = 0$, *and hence* $\int \mathscr{L}(P_t f) \, d\gamma = 0$. *Hence we got the conclusion: Gaussian measure is stationary for the Ornstein–Uhlenbeck process. The proof of 67 is left as homework (use first-order Gaussian integration by parts coordinatewise). It could also be proved by using Green's formula.*

Definition 6.22 (Reversibility of semigroups)**.** A Markov semigroup P_t with stationary measure μ is called reversible if:

$$\int f \cdot P_t g \, d\mu = \int P_t f \cdot g \, d\mu$$

for all $f, g \in L^2(\mu)$.

Reversibility means that, when $X_0 \sim \mu$, the pair (X_0, X_t) has the same law as (X_t, X_0). Indeed,

$$\mathbb{E}[f(X_0)g(X_t)] = \int f P_t g \, d\mu = \int g P_t f \, d\mu = \mathbb{E}[g(X_0)f(X_t)].$$

Equivalently, $P_t f(x) = \mathbb{E}(f(X_0) \mid X_t = x)$ for μ-almost every x: the stationary process has the same transition rule when time is reversed.

Definition 6.23 (Ergodicity)**.** A Markov semigroup P_t with stationary probability measure μ is called ergodic if for all $f \in L^2(\mu)$,

$$P_t f \xrightarrow[t \to \infty]{L_2} \mathbb{E}_\mu f = \int f d\mu$$

If $P_t f$ converges in $L^2(\mu)$ to a constant C, stationarity forces $C = \int f \, d\mu$. Ergodicity asserts that this convergence occurs for every $f \in L^2(\mu)$.

Remark 6.24. *Recall that for every stationary Markov semigroup, the variance* $\mathrm{Var}_\mu(P_t f)$ *is nonincreasing as* $t \to \infty$. *Ergodicity is equivalent to* $\mathrm{Var}_\mu(P_t f) \to 0$ *for every* $f \in L^2(\mu)$.

Definition 6.25 (Dirichlet form of semigroups)**.** Assume X_t is a Markov process with a stationary measure μ and generator $\mathscr{L}$, i.e., $\partial_t(P_t f) = \mathscr{L} P_t f$, and $P_0 f = f$. The Dirichlet form $\mathcal{E}$ is defined by:

$$\mathcal{E}(f, g) := -\int f \mathscr{L} g \, d\mu,$$

initially for g in the generator domain. In the reversible case this form extends to its energy domain; a finite-energy function is one for which $\mathcal{E}(f, f) < \infty$.

Example 6.26 (Dirichlet form of the heat semigroup)**.** *The generator is* $\mathscr{L} = \Delta$ *and the invariant measure is* dx. *Thus*

$$\mathcal{E}(f, g) = -\int f \Delta g \, dx = \int \langle \nabla f, \nabla g \rangle \, dx.$$

Remark 6.27. *The heat semigroup is reversible:* $\int f P_t g \, dx = \int g P_t f \, dx$, *by symmetry of the heat kernel. The details are left as homework.*

Theorem 6.28 (Abstract Theorem about "Poincare inequality")**.** *Let* P_t *be a strongly continuous, reversible and ergodic Markov semigroup with stationary probability measure* μ. *Fix* $c > 0$. *The following are equivalent, with finite-energy functions understood in assertions involving* $\mathcal{E}$*:*

1. *For every finite-energy* f, $\mathrm{Var}_\mu(f) \le c\mathcal{E}(f, f)$ *(Poincare inequality).*
2. *The semigroup has exponential* L^2 *convergence to equilibrium. That is,*

$$\int_{\mathbb{R}^n} \left(P_t f - \int_{\mathbb{R}^n} f d\mu \right)^2 d\mu \le e^{-\frac{2t}{c}} \int_{\mathbb{R}^n} \left(f - \int_{\mathbb{R}^n} f d\mu \right)^2 d\mu$$

   for all $t \ge 0$ *and* $f \in L^2(\mu)$. *This is the same as* $\mathrm{Var}_\mu(P_t f) \le e^{-\frac{2t}{c}} \mathrm{Var}_\mu(f)$.

3. $\mathcal{E}(P_t f, P_t f) \le e^{-\frac{2t}{c}} \mathcal{E}(f, f)$.
4. *For all* $f \in L^2(\mu)$, *there exists a finite constant* $\kappa(f)$ *such that* $\sqrt{\mathrm{Var}_\mu(P_t f)} \le \kappa(f) e^{-\frac{t}{c}}$.
5. *for every finite-energy* f, *there exists a constant* $k(f)$ *such that* $\mathcal{E}(P_t f, P_t f) \le k(f) e^{-\frac{2t}{c}}$

Before we complete the proof, we discuss the most important example in detail, and then the ideas of the proof will become clearer.

6.3 The Ornstein–Uhlenbeck semigroup

Recall $\mathscr{L}u = \Delta u - \langle \nabla u, x\rangle$, where $u\colon \mathbb{R}^n \to \mathbb{R}$ is an appropriate function here. The semigroup P_t is defined as: $\partial_t(P_t f) = \mathscr{L}(P_t f)$, $P_0 f = f$, initially for smooth f in the generator domain.

Lemma 6.29. *A very nice and concrete representation for the Ornstein-Uhlenbeck semigroup:*

1. *Mehler's formula is $P_t f(x) = \mathbb{E}\, f(e^{-t}x + \sqrt{1-e^{-2t}}Z)$, where $Z \sim \mathcal{N}(0, \mathrm{Id})$. It solves $\partial_t P_t f = \mathscr{L} P_t f$ with $P_0 f = f$.*
2. *P_t is ergodic.*
3. *P_t is reversible.*
4. *γ is the stationary measure.*

Claim 6.30. *This claim of second order integration by parts is used in the following proof. For smooth f, g with suitable integrability,*

$$\int g\mathscr{L} f \, d\gamma = -\int \langle \nabla f, \nabla g\rangle \, d\gamma.$$

In particular, the Dirichlet form is given by $\mathcal{E}(f,g) = \int \langle \nabla f, \nabla g\rangle d\gamma$.

Proof of Claim 6.30. By Green's formula, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} g\Delta f \, d\gamma &= c_n \int_{\mathbb{R}^n} \Delta f \cdot g e^{\frac{-|x|^2}{2}} \, dx \\
&= -c_n \int_{\mathbb{R}^n} \left\langle \nabla f, \nabla \left(g e^{\frac{-|x|^2}{2}} \right) \right\rangle dx \\
&= -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle c_n e^{\frac{-|x|^2}{2}} dx + \int_{\mathbb{R}^n} g c_n \langle \nabla f, x\rangle e^{\frac{-|x|^2}{2}} \, dx \\
&= -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle d\gamma + \int_{\mathbb{R}^n} g\langle \nabla f, x\rangle \, d\gamma.
\end{aligned}$$

Therefore,

$$\int_{\mathbb{R}^n} g\mathscr{L} f \, d\gamma = \int_{\mathbb{R}^n} g\Delta f \, d\gamma - \int_{\mathbb{R}^n} g\langle \nabla f, x\rangle \, d\gamma = -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle \, d\gamma.$$

□

Proof of Lemma 6.29. **Proof of Property 4.** First let f be smooth with sufficient integrability. Then,

$$\frac{d}{dt}\int P_t f d\gamma = \int \partial_t \left(P_t f\right) d\gamma = \int \mathscr{L}\left(P_t f\right) d\gamma.$$

Here, if we let $g = 1$, then

$$\int \mathscr{L} f d\gamma = -\int \langle \nabla f, 0 \rangle d\gamma = 0.$$

Thus, $\frac{d}{dt} \int P_t f d\gamma = 0$, implying that $\int P_t f d\gamma = \int f d\gamma$ so the integral of $P_t f$ is constant in time.

Proof of Property 1. We want to show that $P_t f(x) = \mathbb{E}\, f(e^{-t}x + \sqrt{1-e^{-2t}}Z)$ satisfies the following

$$\begin{cases} \partial_t (P_t f) = \mathscr{L}(P_t f) \\ P_0 f = f. \end{cases}$$

By direct calculation, we see that $P_0 f(x) = \mathbb{E}\, f(x) = f(x)$. Next, by the chain rule, we have

$$\begin{aligned} \partial_t (P_t f) &= \partial_t \, \mathbb{E}\, f(e^{-t}x + \sqrt{1-e^{-2t}}Z) \\ &= \mathbb{E}\, \partial_t f(e^{-t}x + \sqrt{1-e^{-2t}}Z) \\ &= \mathbb{E} \left\langle \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}Z), -e^{-t}x + (1-e^{-2t})^{-1/2} e^{-2t} Z \right\rangle \\ &= \int_{\mathbb{R}^n} \left\langle \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}z), -e^{-t}x + (1-e^{-2t})^{-1/2} e^{-2t} z \right\rangle d\gamma(z). \end{aligned}$$

Using first-order Gaussian integration by parts in z, we see that

$$\begin{aligned} &\int_{\mathbb{R}^n} \left\langle \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}z), -e^{-t}x + (1-e^{-2t})^{-1/2} e^{-2t} z \right\rangle d\gamma(z) \\ &= \int_{\mathbb{R}^n} \left\langle \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}z), -e^{-t}x \right\rangle d\gamma(z) + \int_{\mathbb{R}^n} e^{-2t} \Delta f(e^{-t}x + \sqrt{1-e^{-2t}}z)\, d\gamma(z) \\ &= -\left\langle \nabla_x \left[\int_{\mathbb{R}^n} f(e^{-t}x + \sqrt{1-e^{-2t}}z) d\gamma(z) \right], x \right\rangle + \Delta_x \left[\int_{\mathbb{R}^n} f(e^{-t}x + \sqrt{1-e^{-2t}}z)\, d\gamma(z) \right] \\ &= \mathscr{L}\, \mathbb{E}\, f(e^{-t}x + \sqrt{1-e^{-2t}}Z). \end{aligned}$$

Proof of Property 3. Let X, Z be independent standard Gaussian vectors and set $Y = e^{-t}X + \sqrt{1-e^{-2t}}Z$. The pair (X, Y) is centered Gaussian with covariance

$$\begin{pmatrix} \mathrm{Id} & e^{-t}\mathrm{Id} \\ e^{-t}\mathrm{Id} & \mathrm{Id} \end{pmatrix}.$$

Its law is unchanged by exchanging X and Y. Thus

$$\int f P_t g \, d\gamma = \mathbb{E}[f(X)g(Y)] = \mathbb{E}[g(X)f(Y)] = \int g P_t f \, d\gamma.$$

Proof of Property 2. For bounded continuous f, Mehler's formula and bounded convergence give $P_t f(x) \to \mathbb{E}\, f(Z)$ for every x, and then convergence in $L^2(\gamma)$. Density of bounded continuous functions and the contraction property extend this conclusion to every $f \in L^2(\gamma)$. □

Remark 6.31. *For students with a background in stochastic differential equations, the Ornstein–Uhlenbeck process solves $dX_t = -X_t\,dt + \sqrt{2}\,dB_t$, where B_t is standard Brownian motion. Its transition semigroup is the one described above.*

Lemma 6.32 (Gradient commutation and decay)**.** *Let $\{P_t\}$ be the Ornstein-Uhlenbeck semigroup. Then, the following hold:*

1. *$\nabla P_t f = e^{-t} P_t \nabla f$, where $P_t \nabla f = (P_t \partial_1 f, \dots, P_t \partial_n f)$. Here the $\partial_i f$ denote the partial derivatives of f and $\nabla f = (\partial_1 f, \dots, \partial_n f)$.*
2. *$\int |\nabla P_t f|^2 \, d\gamma \le e^{-2t} \int |\nabla f|^2 \, d\gamma$.*

Proof. 1. Recall that $P_t f = \mathbb{E}\, f(e^{-t}x + \sqrt{1-e^{-2t}}Z)$. Then, the chain rule yields that

$$\partial_i (P_t f) = \mathbb{E}\, \partial_{x^i} f\left(e^{-t}x + \sqrt{1-e^{-2t}}Z\right) = e^{-t}\,\mathbb{E}\,\partial_i f\left(e^{-t}x + \sqrt{1-e^{-2t}}Z\right) = e^{-t} P_t(\partial_i f).$$

2. Note that

$$\begin{aligned}\int_{\mathbb{R}^n} |\nabla P_t f|^2 \, d\gamma = e^{-2t} \int_{\mathbb{R}^n} \sum_{1\le i\le n} |P_t \partial_i f|^2 \, d\gamma &= e^{-2t} \sum_{1\le i\le n} \int |P_t \partial_i f|^2 \, d\gamma \\ &\le e^{-2t} \sum_{1\le i\le n} \int_{\mathbb{R}^n} |\partial_i f|^2 \, d\gamma \\ &= e^{-2t} \int_{\mathbb{R}^n} |\nabla f|^2 \, d\gamma,\end{aligned}$$

where we have used $\int |P_t g|^2 d\gamma \le \int g^2 d\gamma$. □

6.4 Gaussian Poincare inequality via the semigroup method

Theorem 6.33 (Gaussian Poincaré Inequality)**.** *Let f be smooth with $f, |\nabla f| \in L^2(\gamma)$. Then,*

$$\mathrm{Var}_\gamma(f) = \int_{\mathbb{R}^n} f^2 d\gamma - \left(\int_{\mathbb{R}^n} f\, d\gamma\right)^2 \le \mathbb{E}_\gamma |\nabla f|^2 = \int_{\mathbb{R}^n} |\nabla f|^2 \, d\gamma,$$

where γ is Gaussian measure.

Proof. Let P_t be the Ornstein-Uhlenbeck semigroup. Then, note that $\int_{\mathbb{R}^n} f^2 d\gamma = \int (P_0 f)^2 d\gamma$ and $\left(\int_{\mathbb{R}^n} f d\gamma\right)^2 = (P_\infty f)^2$ since ergodicity of $\{P_t\}$ implies that $P_\infty f = \int f\, d\gamma$. Since γ is a probability measure, the Fundamental Theorem of Calculus implies that

$$\begin{aligned}\mathrm{Var}_\gamma(f) = \int_{\mathbb{R}^n} (P_0 f)^2 d\gamma - \int_{\mathbb{R}^n} (P_\infty f)^2 \, d\gamma &= \int_{\mathbb{R}^n} (P_0 f)^2 - (P_\infty f)^2 \, d\gamma \\ &= -\int_{\mathbb{R}^n} \int_0^\infty \frac{\partial}{\partial t} (P_t f)^2 \, dt \, d\gamma(x) \\ &= -\int_0^\infty \int_{\mathbb{R}^n} \frac{\partial}{\partial t} (P_t f)^2 \, d\gamma(x) \, dt \\ &= -\int_0^\infty \int_{\mathbb{R}^n} 2 P_t f \cdot \mathscr{L}(P_t f) \, d\gamma(x) \, dt\end{aligned}$$

Applying Gaussian integration by parts, we obtain

$$\begin{aligned}
-2\int_0^\infty \int_{\mathbb{R}^n} P_t f \cdot \mathscr{L}(P_t f)\, d\gamma(x)\, dt &= 2\int_0^\infty \int_{\mathbb{R}^n} \langle \nabla P_t f, \nabla P_t f\rangle \; d\gamma\, dt \\
&= 2\int_0^\infty \int_{\mathbb{R}^n} e^{-2t} \left|P_t \nabla f\right|^2 \; d\gamma\, dt \\
&\leq 2\int_0^\infty \int_{\mathbb{R}^n} e^{-2t} \left|\nabla f\right|^2 \; d\gamma\, dt \\
&= \int_{\mathbb{R}^n} \left|\nabla f\right|^2 \; d\gamma.
\end{aligned}$$

□

Remark 6.34. *When is the Gaussian Poincaré inequality sharp? The only inequality in the proof is L^2 contraction applied to each $\partial_i f$. Equality holds precisely when all these derivatives are constant, so f is affine: $f(x) = a + \langle x, \theta\rangle$. Indeed,*

$$\mathrm{Var}_\gamma(a + \langle x, \theta\rangle) = \sum_i \theta_i^2 \,\mathrm{Var}_\gamma(x_i) = |\theta|^2 = \int |\nabla f|^2 \, d\gamma.$$

6.5 A discussion on Poincare inequalities, the example of the circle and periodic functions

More generally, let $d\mu = e^{-V(x)}\, dx$ for $V \in \mathcal{C}^2(\mathbb{R}^n)$ such that $\int d\mu = 1$. Then we define the operator $\mathscr{L}_\mu u = \Delta u - \langle \nabla V, \nabla u\rangle$. In the Gaussian case $d\gamma = (2\pi)^{-n/2} e^{-|x|^2/2}\, dx$, we have $V(x) = |x|^2/2 + (n/2)\log(2\pi)$, $\nabla V = x$ and $\mathscr{L}_\gamma u = \Delta u - \langle x, \nabla u\rangle$, which is exactly the Ornstein-Uhlenbeck operator. For a given measure μ we have the integration by parts formula

$$\int f \cdot \mathscr{L}_\mu g \, d\mu = -\int \langle \nabla f, \nabla g\rangle \, d\mu.$$

Definition 6.35. The *spectral gap* of $-\mathscr{L}_\mu$ is

$$\lambda = \inf_{\substack{\int f\, d\mu = 0 \\ \int f^2\, d\mu > 0}} \frac{\int |\nabla f|^2 \, d\mu}{\int f^2 \, d\mu}.$$

If the infimum is attained by a nonzero function f, it is called a *first eigenfunction*, and

$$\mathscr{L}_\mu f = -\lambda f. \tag{$\star$}$$

When $\lambda > 0$ and is attained, it is the first positive eigenvalue of $-\mathscr{L}_\mu$ (equivalently, $-\lambda$ is the first nonzero eigenvalue of $\mathscr{L}_\mu$).

Definition 6.36. The Poincaré constant $c_p(\mu)$ is the smallest constant, possibly $+\infty$, such that

$$\mathrm{Var}_\mu(f) \le c_p(\mu) \int |\nabla f|^2 \, d\mu$$

for every reasonable function f.

Remark 6.37. *The Gaussian Poincaré inequality and equality for linear functions imply* $c_p(\gamma) = 1$.

Claim 6.38. *For a probability measure* μ, $c_p(\mu) = 1/\lambda$, *with* $1/0 = +\infty$.

Remark 6.39. *This identity follows directly from the variational definition and does not require an eigenfunction to exist. For example, for the two-sided exponential probability measure* $\frac{1}{2}e^{-|t|}\,dt$, *the spectral gap is not attained. Lebesgue measure is an infinite-measure example and is outside the probability-measure formulation above. For Gaussian measure, the centered linear functions are first eigenfunctions of the Ornstein–Uhlenbeck operator.*

It is also possible to use harmonic analysis to prove the Gaussian Poincaré inequality. We will illustrate this with a simpler example. Consider 2π-periodic functions with two continuous derivatives which are themselves also 2π-periodic.

Theorem 6.40. *Let* f *be a* 2π*-periodic function with two continuous and* 2π*-periodic derivatives. Then*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f^2(\theta)\,d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\,d\theta\right)^2 \le \frac{1}{2\pi}\int_{-\pi}^{\pi} \dot f(\theta)^2\,d\theta,$$

where $\dot f$ *is the derivative of* f.

Proof. Use the 2π-periodic Fourier convention

$$f(\theta) = \sum_{k\in\mathbb{Z}} \widehat f(k) e^{ik\theta}, \qquad \widehat f(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta) e^{-ik\theta}\,d\theta.$$

Plancherel gives

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f^2\,d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f\,d\theta\right)^2 = \sum_{k\neq 0} |\widehat f(k)|^2.$$

Since $\widehat{\dot f}(k) = ik\widehat f(k)$, another application gives

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |\dot f|^2\,d\theta = \sum_{k\in\mathbb{Z}} k^2 |\widehat f(k)|^2 \ge \sum_{k\neq 0} |\widehat f(k)|^2.$$

□

Remark 6.41. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is* 1*-Lipschitz, then* $|\nabla f| \le 1$ *almost everywhere, hence* $\mathrm{Var}_\gamma(f) \le 1$. *More generally,* $\mathrm{Var}_\gamma(f) \le L^2$ *for an* L*-Lipschitz function.*

6.6 Proof of the abstract Poincare-type inequality for stationary measures of semigroups

Now we go back to our general Markov process and see how much of the previous theory carries over.

Theorem 6.42. *Let P_t be a strongly continuous, reversible, ergodic Markov semigroup with stationary probability measure μ. The following are equivalent for a given constant $c > 0$.*

1. $\mathrm{Var}_\mu(f) \le c \cdot \mathcal{E}(f,f)$ *for every finite-energy f,*
2. $\left\|P_t f - \int (P_t f)\, d\mu\right\|_{L^2(\mu)} \le e^{-t/c} \left\|f - \int f\, d\mu\right\|_{L^2(\mu)}$ *for all $t \ge 0$ and $f \in L^2(\mu)$,*
3. $\mathcal{E}(P_t f, P_t f) \le e^{-2t/c}\mathcal{E}(f,f)$ *for all $t \ge 0$ and every finite-energy f,*
4. *for all $f \in L^2(\mu)$ there is a constant $\kappa(f)$ such that $\left\|P_t f - \int (P_t f)\, d\mu\right\|_{L^2(\mu)} \le \kappa(f) \cdot e^{-t/c}$ for all $t \ge 0$,*
5. *for every finite-energy f there is a finite constant $\kappa(f)$ such that $\mathcal{E}(P_t f, P_t f) \le \kappa(f) \cdot e^{-2t/c}$ for all $t \ge 0$.*

So far, for the Gaussian measure, we have used the implication 3. $\Rightarrow$ 1 to prove the Gaussian Poincaré inequality (Theorem 6.33). It is important to note that the above result is not strictly speaking a generalization of the Poincaré inequality. It establishes the equivalence of several conditions, and we had to do work by hand to prove energy decay (condition 3.) in the case of the Gaussian measure.

Lemma 6.43. *Let P_t be a semigroup satisfying the conditions of Theorem 6.42. Then*

$$\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\mathcal{E}(P_t f, P_t f)$$

for all reasonable functions f.

Proof. Recall

$$\mathrm{Var}_\mu(P_t f) = \int (P_t f)^2\, d\mu - \left(\int P_t f\, d\mu\right)^2.$$

By stationarity, the second term is constant, so

$$\begin{aligned}
\frac{d}{dt}\mathrm{Var}_\mu(P_t f) &= \frac{d}{dt}\int (P_t f)^2\, d\mu \\
&= \int \frac{d}{dt}(P_t f)^2\, d\mu \\
&= 2\int (P_t f)\frac{d}{dt}(P_t f)\, d\mu \\
&= 2\int P_t f \mathscr{L}(P_t f)\, d\mu \\
&= -2\mathcal{E}(P_t f, P_t f).
\end{aligned}$$

□

Remark 6.44. *Recall that* $\mathrm{Var}_\mu(P_t f)$ *decreases to zero as* $t \to \infty$ *and so the above implies that* $\mathcal{E}(g,g) \geq 0$ *for all reasonable* g*.*

Corollary 6.45 (Integral representation of variance)**.** *Let* $\{P_t\}$ *be a semigroup satisfying the conditions of Theorem 6.42. Then*

$$\mathrm{Var}_\mu(f) = 2\int_0^\infty \mathcal{E}(P_t f, P_t f)\, dt.$$

Proof. Integrating the derivative identity from 0 to T gives

$$\mathrm{Var}_\mu(f) - \mathrm{Var}_\mu(P_T f) = 2\int_0^T \mathcal{E}(P_t f, P_t f)\, dt.$$

Ergodicity gives $\mathrm{Var}_\mu(P_T f) \to 0$ as $T \to \infty$, proving the formula. □

With these lemmas, we can prove some of the implications required for Theorem 6.42. For now we will not use the reversibility assumption.

Proof of Theorem 6.42 Part 1. In this part of the proof we will show

$$5 \Leftarrow 3 \Rightarrow 1 \Leftrightarrow 2 \Rightarrow 4,$$

without using reversibility.
(2 ⇒ 4): This implication is immediate by setting $\kappa(f) = \left\| f - \int f\, d\mu \right\|_{L^2(\mu)}$.
(3 ⇒ 5): This time setting $\kappa(f) = \mathcal{E}(f,f)$ is sufficient.
(3 ⇒ 1): We can make a direct calculation using Corollary 6.45

$$\begin{aligned}\mathrm{Var}_\mu(f) &= 2\int_0^\infty \mathcal{E}(P_t f, P_t f)\, dt \\ &\leq 2\int_0^\infty e^{-2t/c}\mathcal{E}(f,f)\, dt \\ &= c\mathcal{E}(f,f).\end{aligned}$$

(1 ⇒ 2): Suppose that $\mathrm{Var}_\mu(f) \leq c\mathcal{E}(f,f)$ for all reasonable functions f. Then, by Lemma 6.43

$$\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\mathcal{E}(P_t f, P_t f) \leq -\frac{2}{c}\mathrm{Var}_\mu(P_t f).$$

We can now apply Grönwall's inequality viewing $\mathrm{Var}_\mu(P_t f)$ as a function of t to deduce that

$$\mathrm{Var}_\mu(P_t f) \leq e^{-2t/c}\mathrm{Var}_\mu(f).$$

By taking the square root, we see that

$$\left\| P_t f - \int P_t f\, d\mu \right\|_{L^2(\mu)} \leq e^{-t/c} \left\| f - \int f\, d\mu \right\|_{L^2(\mu)}.$$

($2 \Rightarrow 1$): We will use Lemma 6.43 in reverse and expand the definition of derivative.

$$\begin{aligned} 2\mathcal{E}(f,f) &= \lim_{t\to 0} \frac{\operatorname{Var}_\mu(f) - \operatorname{Var}_\mu(P_t f)}{t} \\ &\geq \operatorname{Var}_\mu(f) \cdot \lim_{t \downarrow 0} \frac{1 - e^{-2t/c}}{t} = \frac{2}{c} \operatorname{Var}_\mu(f), \end{aligned}$$

where the first inequality follows from the assumption 2.. Rearranging the constants, 1 follows. □

To prove the implication $1 \Rightarrow 2$ we used Grönwall's inequality.

Lemma 6.46 (Grönwall's inequality)**.** *Let $g : [0,\infty) \to \mathbb{R}$ be differentiable. If*

$$\frac{d}{dt} g(t) \leq c \cdot g(t),$$

then

$$g(t) \leq e^{ct} g(0).$$

Proof. Multiplying by the integrating factor gives

$$\frac{d}{dt}\big(e^{-ct} g(t)\big) = e^{-ct}\big(g'(t) - c g(t)\big) \leq 0.$$

Thus $e^{-ct} g(t) \leq g(0)$ for $t \geq 0$. This argument also covers $g(0) = 0$ and does not require taking a logarithm. □

Next, we will suppose that the chain P_t is reversible.

Lemma 6.47. *If $\{P_t\}$ is a reversible Markov semigroup, then, wherever the respective quantities are positive,*

- $\log \operatorname{Var}_\mu(P_t f)$ *is convex in $t \geq 0$, and*
- $\log \mathcal{E}(P_t f, P_t f)$ *is convex in $t \geq 0$.*

Recall that g is called convex if for all t, s and all $\lambda \in [0,1]$

$$g(\lambda t + (1-\lambda)s) \leq \lambda g(t) + (1-\lambda) g(s),$$

and that if $g \in \mathcal{C}^2$ then this is equivalent to $g'' \geq 0$.

Proof. Subtract the mean of f, which changes neither variance nor energy, and put $u = P_t f$. Then $\operatorname{Var}_\mu(u) = \int u^2 \, d\mu$. Differentiation gives

$$\frac{d^2}{dt^2} \log \operatorname{Var}_\mu(u) = -\frac{2\frac{d}{dt}\mathcal{E}(u,u)}{\operatorname{Var}_\mu(u)} - \frac{4\mathcal{E}(u,u)^2}{\operatorname{Var}_\mu(u)^2}. \tag{68}$$

Reversibility makes $\mathcal{E}$ symmetric, and hence

$$\frac{d}{dt}\mathcal{E}(u,u) = 2\mathcal{E}(\mathscr{L}u,u) = -2\int(\mathscr{L}u)^2\,d\mu. \tag{69}$$

Therefore,

$$\frac{d^2}{dt^2}\log \mathrm{Var}_\mu(u) = \frac{4\left(\int u^2\,d\mu\int(\mathscr{L}u)^2\,d\mu - (\int u\mathscr{L}u\,d\mu)^2\right)}{\mathrm{Var}_\mu(u)^2} \geq 0$$

by Cauchy–Schwarz.

For energy, write $e(t) = \mathcal{E}(u,u)$. A further differentiation yields $e''(t) = 4\mathcal{E}(\mathscr{L}u,\mathscr{L}u)$, while $e'(t) = 2\mathcal{E}(u,\mathscr{L}u)$. The Cauchy–Schwarz inequality for the nonnegative symmetric form $\mathcal{E}$ gives

$$e'(t)^2 \leq 4\mathcal{E}(u,u)\mathcal{E}(\mathscr{L}u,\mathscr{L}u) = e(t)e''(t).$$

This proves the convexity of $\log e(t)$. If variance or energy vanishes identically, its decay assertions are immediate. □

Proof of Theorem 6.42, Part 2. It remains to use reversibility to prove $1 \Rightarrow 3$, $4 \Rightarrow 2$, and $5 \Rightarrow 3$.

For $1 \Rightarrow 3$, put $u = P_tf - \int f\,d\mu$. By Cauchy–Schwarz and condition 1,

$$\mathcal{E}(u,u)^2 \leq \mathrm{Var}_\mu(u)\int(\mathscr{L}u)^2\,d\mu \leq c\mathcal{E}(u,u)\int(\mathscr{L}u)^2\,d\mu.$$

Together with (69), this gives $\frac{d}{dt}\mathcal{E}(P_tf,P_tf) \leq -(2/c)\mathcal{E}(P_tf,P_tf)$. Grönwall's inequality proves 3.

For the other two implications, use the following elementary consequence of log-convexity. If $g(t) > 0$ is log-convex and $g(t) \leq Ke^{-at}$, then for $0 < t < T$,

$$\log g(t) \leq \left(1 - \frac{t}{T}\right)\log g(0) + \frac{t}{T}(\log K - aT).$$

Letting $T \to \infty$ gives $g(t) \leq e^{-at}g(0)$. Apply this first to $g(t) = \mathrm{Var}_\mu(P_tf)$, squaring the estimate in condition 4, and then to $g(t) = \mathcal{E}(P_tf,P_tf)$ using condition 5. Zero cases follow directly, or by continuity. This completes all the equivalences. □

6.7 The Gaussian Log-Sobolev Inequality of Gross via the semigroup method

In this section, we will establish the log Sobolev inequality via properties of the Ornstein-Uhlenbeck semigroup. To start, we recall some basic properties of this semigroup:

- for all $t \geq 0$, $P_tf(x) = \mathbb{E}f(e^{-t}x + \sqrt{1-e^{-2t}}Z)$ where $Z \sim \mathcal{N}(0,I_n)$; in particular, $P_0f = f$,

- $d\gamma = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{|x|^2}{2}\right) dx$ — the Gaussian probability measure is stationary,
- $\frac{d}{dt} P_t f = \mathscr{L} P_t f$, where
$$\mathscr{L} u = \Delta u - \langle x, \nabla u \rangle.$$
In particular,
$$\int_{\mathbb{R}^n} f \mathscr{L} g \, d\gamma = -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g \rangle \, d\gamma.$$

Furthermore, note that this process is reversible and ergodic:

$$\lim_{t\to\infty} P_t f = \int_{\mathbb{R}^n} f \, d\gamma = \mathbb{E}(f(Z)).$$

Theorem 6.48 (Gross's Log-Sobolev Inequality)**.** *Let $f \geq 0$ be smooth, with the integrals below finite. Then*

$$\int_{\mathbb{R}^n} f \log f \, d\gamma - \int_{\mathbb{R}^n} f \, d\gamma \cdot \log\left(\int_{\mathbb{R}^n} f \, d\gamma\right) \leq \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \, d\gamma$$

To establish this inequality, we will reproduce the proof of the Gaussian-Poincaré inequality in a more general setting. Let ϕ be a convex C^2 function on an interval containing the range of f. As in the Gaussian Poincaré proof,

$$\begin{aligned}
\int \phi(f) \, d\gamma - \phi\left(\int f \, d\gamma\right) &= -\int_0^\infty \int \phi'(P_t f) \mathscr{L}(P_t f) \, d\gamma \, dt \\
&= \int_0^\infty \int \phi''(P_t f) |\nabla P_t f|^2 \, d\gamma \, dt \\
&= \int_0^\infty e^{-2t} \left[\int \phi''(P_t f) |P_t \nabla f|^2 \, d\gamma\right] dt.
\end{aligned}$$

Here the second equality is Gaussian integration by parts and the last one uses gradient commutation. For $\phi(s) = s \log s$, first take f bounded above and bounded away from zero; adding a small positive constant and truncating extends the result to nonnegative f. We use the convention $0 \log 0 = 0$ and interpret the energy by this limiting procedure.

Proof of Theorem 6.48. Taking $\phi(t) = t \log t$ in the computation above, we see that

$$\int_{\mathbb{R}^n} f \log f \, d\gamma - \int_{\mathbb{R}^n} f \, d\gamma \cdot \log\left(\int_{\mathbb{R}^n} f \, \mathrm{d}\gamma\right) = \int_0^\infty e^{-2t} \left[\int_{\mathbb{R}^n} \frac{|P_t \nabla f|^2}{P_t f} \, d\gamma\right] dt.$$

Using the explicit representation for P_t and applying Cauchy-Schwarz, we see that for any functions h, g

$$[P_t(hg)]^2 \leq P_t(h^2) \cdot P_t(g^2).$$

Therefore, we see that

$$\begin{aligned}
\int_{\mathbb{R}^n} \frac{1}{P_t f} \cdot |P_t \nabla f|^2 \, d\gamma &= \int_{\mathbb{R}^n} \frac{1}{P_t f} \cdot \sum_{1 \le i \le n} (P_t \partial_i f)^2 \, d\gamma \\
&= \int_{\mathbb{R}^n} \frac{1}{P_t f} \cdot \sum_{1 \le i \le n} \left[P_t \left(\frac{\partial_i f}{\sqrt{f}} \cdot \sqrt{f} \right) \right]^2 d\gamma \\
&\le \int_{\mathbb{R}^n} \frac{1}{P_t f} \cdot \sum_{1 \le i \le n} P_t \left(\frac{(\partial_i f)^2}{f} \right) \cdot P_t(f) \, d\gamma \\
&= \int_{\mathbb{R}^n} P_t \left(\frac{|\nabla f|^2}{f} \right) d\gamma \\
&= \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \, d\gamma,
\end{aligned}$$

where the last equality follows from stationarity. Integrating e^{-2t} over $[0, \infty)$ gives the factor $1/2$ and proves the theorem. □

Remark 6.49. *For a probability measure μ and a nonnegative integrable function f, we define its* ***entropy*** *with respect to μ by*

$$\mathrm{Ent}_\mu(f) = \int f \log f \, \mathrm{d}\mu - \left(\int f \, d\mu \right) \cdot \log \left(\int f \, d\mu \right)$$

By applying Jensen's inequality to $t \mapsto t \log t$, we see that $\mathrm{Ent}_\mu(\cdot) \ge 0$.

Remark 6.50. *The quantity $\int |\nabla f|^2 / f \, d\mu$ is commonly referred to as the* ***Fisher information****.*

Remark 6.51. *If we let $f = g^2$, then*

$$\int_{\mathbb{R}^n} g^2 \log(g^2) \, d\gamma - \int_{\mathbb{R}^n} g^2 \, d\gamma \cdot \log \left(\int_{\mathbb{R}^n} g^2 \, d\gamma \right) \le 2 \int_{\mathbb{R}^n} |\nabla g|^2 \, d\gamma$$

Remark 6.52. *See the HW file for a discussion of other related inequalities.*

6.8 Gaussian isoperimetry via the semigroup method

Definition 6.53. Let $A \subseteq \mathbb{R}^n$ be a Borel measurable set. We define the **Minkowski content** of $\partial A \subseteq \mathbb{R}^n$ by

$$|\partial A| = \liminf_{\varepsilon \searrow 0} \frac{|(A + \varepsilon \mathbf{B}_2^n) \setminus A|}{\varepsilon}.$$

The classical isoperimetric inequality states the following:

Theorem 6.54 (Classical Isoperimetric Inequality)**.** *Let $A \subseteq \mathbb{R}^n$ be a Borel measurable set such that $|A| = |\mathbf{B}_2^n|$. Then, $|\partial A| \geq |\partial \mathbf{B}_2^n|$.*

Remark 6.55. *By scaling, we see that if $|A| = |R\mathbf{B}_2^n|$, then*

$$|\partial A| \geq |\partial (R\mathbf{B}_2^n)|.$$

Definition 6.56. For a Borel measurable set $A \subseteq \mathbb{R}^n$, we define its **Gaussian perimeter** by

$$\gamma^+(\partial A) = \liminf_{\varepsilon \searrow 0} \frac{\gamma((A + \varepsilon \mathbf{B}_2^n) \setminus A)}{\varepsilon},$$

where $\gamma(\cdot)$ denotes the standard Gaussian measure on $\mathbb{R}^n$.

Remark 6.57. *If A is a "nice enough" set, then*

$$|\partial A| = \int_{\partial A} \mathbb{1}\, d\mathcal{H}^{n-1},$$

where $\mathcal{H}^{n-1}$ denotes the $(n-1)$-dimensional Hausdorff measure. Similarly,

$$\gamma^+(\partial A) = \int_{\partial A} \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}\, d\mathcal{H}^{n-1}(x).$$

Heuristically, we also have

$$\gamma^+(\partial A) \approx \frac{1}{\varepsilon} \mathbb{P}\{Z \notin A : \operatorname{dist}(Z, A) \leq \varepsilon\},$$

where Z is a random variable on $\mathbb{R}^n$ with law γ.

Theorem 6.58 (Gaussian Isoperimetric Inequality, Tsirelson and Sudakov (1974), Borell (1975))**.** *Let $A \subseteq \mathbb{R}^n$ be Borel measurable, with $a = \gamma(A) \in (0,1)$. Let $H = \{x : \langle x, \theta \rangle \leq \Phi^{-1}(a)\}$, where $\theta \in \mathbb{S}^{n-1}$. Then $\gamma(H) = \gamma(A)$ and $\gamma^+(\partial A) \geq \gamma^+(\partial H)$. For $a = 0$ or $a = 1$, the corresponding lower bound is zero.*

Claim 6.59. *Let $a \in (0,1)$, $\theta \in \mathbb{S}^{n-1}$, and $H = \{x : \langle x, \theta \rangle \leq \Phi^{-1}(a)\}$, where $\Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-t^2/2}\, dt$ denotes the CDF of the standard Gaussian distribution. Then, $\gamma(H) = a$.*

Proof. By an orthogonal change of variables, assume $\theta = e_1$. Integrating out the remaining Gaussian coordinates gives

$$\gamma(H) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Phi^{-1}(a)} e^{-x_1^2/2}\, dx_1 = a.$$

□

Remark 6.60. *This claim implies that* $\gamma^+(\partial H) = \frac{1}{\sqrt{2\pi}} e^{-\Phi^{-1}(a)^2/2}$. *Therefore, the Gaussian isoperimetric inequality implies that*

$$\gamma^+(\partial A) \geq \frac{1}{\sqrt{2\pi}} e^{-\Phi^{-1}(a)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-\Phi^{-1}(\gamma(A))^2/2}.$$

Definition 6.61. Let $a \in (0,1)$. We define the **Gaussian isoperimetric profile** as

$$I(a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\Phi^{-1}(a)^2}{2}} = \inf_{\substack{\gamma(A)=a, \\ A \subseteq \mathbb{R}^n}} \gamma^+(\partial A).$$

Remark 6.62. *Here are some basic properties of* $I(\cdot)$*:*

- *The function* $a \mapsto I(a)$ *is concave, with continuous endpoint values* $I(0) = I(1) = 0$.
- *The function* $s \mapsto I(1/2 + s)$ *is even on* $[-1/2, 1/2]$.
- I *is increasing on* $[0, 1/2]$ *and decreasing on* $[1/2, 1]$.
- $I''(a)I(a) = -1$ *for* $0 < a < 1$.

6.9 Bobkov's Inequality via the semigroup method

This section is closely related to the Gaussian isoperimetry, specifically the Gaussian Isoperimetric Inequality and the Gaussian isoperimetric profile in the last section. You may review them before proceeding to this section.

Theorem 6.63 (Bobkov's Inequality - "functional version of the Gaussian Isoperimetric Inequality")**.** *Let* $f \in C^1(\mathbb{R}^n)$ *take values in* $[0,1]$*, and assume the integral on the right is finite. Then*

$$I\left(\int_{\mathbb{R}^n} f \, d\gamma\right) \leq \int_{\mathbb{R}^n} \sqrt{|\nabla f|^2 + I(f)^2} \, d\gamma$$

We use the continuous endpoint values $I(0) = I(1) = 0$.

Claim 6.64. *Bobkov's inequality implies the Gaussian isoperimetric inequality for sets with smooth boundary.*

Proof sketch. First let A be closed with smooth boundary and finite Gaussian perimeter. Approximate $\mathbb{1}_A$ by the Lipschitz functions

$$f_\varepsilon(x) = \max\{1 - \operatorname{dist}(x, A)/\varepsilon, 0\}.$$

Smoothing allows us to apply Bobkov's inequality. Since $f_\varepsilon \to \mathbb{1}_A$ almost everywhere and $I(0) = I(1) = 0$,

$$\int f_\varepsilon \, d\gamma \to \gamma(A), \qquad \int I(f_\varepsilon) \, d\gamma \to 0.$$

Moreover, $|\nabla f_\varepsilon| \le \varepsilon^{-1}\mathbb{1}_{A_\varepsilon \setminus A}$ almost everywhere. Hence

$$I\left(\int f_\varepsilon \, d\gamma\right) \le \int \sqrt{I(f_\varepsilon)^2 + |\nabla f_\varepsilon|^2}\, d\gamma \le \int I(f_\varepsilon)\, d\gamma + \frac{\gamma(A_\varepsilon) - \gamma(A)}{\varepsilon}.$$

Taking the lower limit gives $I(\gamma(A)) \le \gamma^+(\partial A)$. No connectedness assumption is needed. Standard approximation extends the statement to Borel sets; if their outer Minkowski content is infinite, the claim is immediate. □

6.9.1 Semigroup proof of Bobkov's inequality

Let $\{P_t\}_{t\ge 0}$ be the Ornstein-Uhlenbeck semigroup, then positivity and preservation of constants imply that

$$f \in [0,1] \Rightarrow P_t f \in [0,1].$$

Recall that

$$P_0 f = f,\ P_\infty f = \int f \, d\gamma$$

and with $\mathscr{L}$ being Ornstein-Uhlenbeck operator

$$\frac{d}{dt} P_t f = \mathscr{L}(P_t f) \text{ s.t. } \int f \mathscr{L} g \, d\gamma = -\int \langle \nabla f, \nabla g \rangle \, d\gamma$$

Proof idea: We will show $\frac{d}{dt}\int \sqrt{|\nabla(P_t f)|^2 + I(P_t f)^2}\, d\gamma \le 0$.

This implies that $\int \sqrt{|\nabla(P_t f)|^2 + I(P_t f)^2}\, d\gamma$ is nonincreasing in $t \ge 0$. Thus,

$$\int \sqrt{|\nabla(P_0 f)|^2 + I(P_0 f)^2}\, d\gamma \ge \int \sqrt{|\nabla(P_\infty f)|^2 + I(P_\infty f)^2}\, d\gamma$$

Since $P_\infty f = \int f\, d\gamma$, $\nabla P_\infty f = \nabla(\text{constant}) = 0$. Then,

$$\int \sqrt{|\nabla f|^2 + I(f)^2}\, d\gamma \ge \int \sqrt{I\left(\int f\, d\gamma\right)^2}\, d\gamma = I\left(\int f\, d\gamma\right)$$

which implies Bobkov's inequality. For bounded derivatives, gradient commutation gives $\|\nabla P_t f\|_\infty \le e^{-t}\|\nabla f\|_\infty \to 0$; Mehler's formula and dominated convergence justify the displayed limit of the functional.

It remains to prove $\frac{d}{dt}\int \sqrt{|\nabla(P_t f)|^2 + I(P_t f)^2}\, d\gamma \le 0$. By the semigroup property, it is enough to verify this at $t = 0$ for an arbitrary starting function. For the calculation, first assume f is smooth with bounded derivatives and takes values in $[\eta, 1-\eta]$ for some $\eta > 0$. Smoothing and then letting $\eta \downarrow 0$ gives the stated result.

Setup and Preliminary Identities: The following identities will be used in the proof of Bobkov's inequality:

1. Define $h(x) = I(f(x))^2 + |\nabla f(x)|^2$. Then, chain rule implies that $\nabla h = 2I(f)I'(f)\nabla f + 2\nabla^2 f \cdot \nabla f$.

2. $\nabla(\mathscr{L} f) = \mathscr{L}\nabla f - \nabla f$. Indeed,

$$\nabla(\mathscr{L} f) = \nabla(\Delta f - \langle \nabla f, x\rangle) = \nabla\Delta f - \nabla^2 f x - \nabla f = \mathscr{L}\nabla f - \nabla f.$$

   Write $\mathscr{L}\nabla f = (\mathscr{L}\partial_1 f, \ldots, \mathscr{L}\partial_n f)$.

3. $I \cdot I'' = -1$. To verify this, note that

$$\begin{aligned} &I(a) = \frac{1}{\sqrt{2\pi}} e^{-\Phi^{-1}(a)^2/2};\ I'(a) = \frac{-1}{\sqrt{2\pi}} \Phi^{-1}(a)' \cdot \Phi^{-1}(a) e^{-\Phi^{-1}(a)^2/2} \\ &\left(\Phi^{-1}\right)' = \frac{1}{\Phi'\left(\Phi^{-1}(a)\right)} = \sqrt{2\pi} e^{\Phi^{-1}(a)^2/2} = \frac{1}{I(a)} \\ &\Rightarrow I'(a) = -\Phi^{-1}(a);\ I''(a) = -\frac{1}{I(a)} \end{aligned}$$

4. Item 3 also implies the identity: $(I(a)I'(a))' = (I')^2 - 1$.

Proof of Theorem 6.63. Recall $h = I(f)^2 + |\nabla f|^2$. Differentiation and the commutation identity give

$$\begin{aligned} D &:= \frac{d}{dt} \int \sqrt{I(P_t f)^2 + |\nabla P_t f|^2}\, d\gamma \Big|_{t=0} \\ &= \int \frac{I(f)I'(f)\mathscr{L} f + \langle \nabla f, \mathscr{L}\nabla f\rangle - |\nabla f|^2}{\sqrt{h}}\, d\gamma. \end{aligned}$$

We integrate the first term by parts:

$$\begin{aligned} \int \frac{I(f)I'(f)\mathscr{L} f}{\sqrt{h}}\, d\gamma &= -\int \left\langle \nabla f, \nabla\left(\frac{I(f)I'(f)}{\sqrt{h}}\right)\right\rangle d\gamma \\ &= -\int \frac{(I'(f)^2 - 1)|\nabla f|^2}{\sqrt{h}}\, d\gamma + \int \frac{I(f)I'(f)\langle \nabla f, \nabla h\rangle}{2h^{3/2}}\, d\gamma. \end{aligned}$$

Its positive $|\nabla f|^2/\sqrt{h}$ term cancels the last term in D. For the vector term, coordinatewise integration by parts gives

$$\begin{aligned} \int \frac{\langle \nabla f, \mathscr{L}\nabla f\rangle}{\sqrt{h}}\, d\gamma &= -\sum_i \int \left\langle \nabla\partial_i f, \frac{\nabla\partial_i f}{\sqrt{h}} - \frac{\partial_i f \nabla h}{2h^{3/2}}\right\rangle d\gamma \\ &= -\int \frac{\|\nabla^2 f\|_{HS}^2}{\sqrt{h}}\, d\gamma + \int \frac{\langle \nabla^2 f \nabla f, \nabla h\rangle}{2h^{3/2}}\, d\gamma. \end{aligned}$$

Using $\nabla h = 2I(f)I'(f)\nabla f + 2\nabla^2 f \nabla f$, we obtain

$$D = \int \frac{-h\big(I'(f)^2|\nabla f|^2 + \|\nabla^2 f\|_{HS}^2\big) + |I(f)I'(f)\nabla f + \nabla^2 f \nabla f|^2}{h^{3/2}}\, d\gamma = -\int \frac{M}{h^{3/2}}\, d\gamma,$$

where expansion gives

$$\begin{aligned} M &= I(f)^2\|\nabla^2 f\|_{HS}^2 + I'(f)^2|\nabla f|^4 - 2I(f)I'(f)\langle \nabla^2 f \nabla f, \nabla f\rangle \\ &\quad + |\nabla f|^2\|\nabla^2 f\|_{HS}^2 - |\nabla^2 f \nabla f|^2 \\ &= \|I(f)\nabla^2 f - I'(f)\nabla f \otimes \nabla f\|_{HS}^2 + |\nabla f|^2\|\nabla^2 f\|_{HS}^2 - |\nabla^2 f \nabla f|^2 \geq 0. \end{aligned}$$

The last inequality uses $|Av|^2 \leq \|A\|_{HS}^2|v|^2$. This argument works for either sign of $I'(f)$. Thus $D \leq 0$, and the semigroup monotonicity established above proves Bobkov's inequality. □

6.9.2 Applications to concentration of measure

Definition 6.65. Let $A \subseteq \mathbb{R}^n$ be a Borel measurable set and $t > 0$. Define

$$A_t := \{x \in \mathbb{R}^n : \operatorname{dist}(x, A) \leq t\}.$$

Theorem 6.66 (Gaussian isoperimetry in enlargement form)**.** *Let $A \subseteq \mathbb{R}^n$ be Borel measurable with $0 < \gamma(A) < 1$. If H is a half-space with $\gamma(H) = \gamma(A)$, then for $t \geq 0$,*

$$\gamma(A_t) \geq \gamma(H_t) = \Phi\big(\Phi^{-1}(\gamma(A)) + t\big).$$

In particular, if $\gamma(A) \geq 1/2$, then $\gamma(A_t) \geq 1 - \frac{1}{2}e^{-t^2/2}$.

Remark 6.67. *Indeed, $\gamma(A) = \gamma(H)$, $H_t = \{x : \langle x, \theta\rangle \leq \Phi^{-1}(\gamma(A)) + t\}$*

$$\gamma\left(H_t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Phi^{-1}(\gamma(A))+t} e^{-s^2/2}\, ds = \Phi(\Phi^{-1}(\gamma(A)) + t)$$

Proof. Idea: Write A_t in layers. For a smooth set, at times when differentiation is valid, let $h(t) = \Phi^{-1}(\gamma(A_t))$. Gaussian isoperimetry gives

$$h'(t) = \frac{\frac{d}{dt}\gamma(A_t)}{I(\gamma(A_t))} = \frac{\gamma^+(\partial A_t)}{I(\gamma(A_t))} \geq 1.$$

Integrating yields $h(t) \geq h(0) + t$, hence $\gamma(A_t) \geq \Phi(\Phi^{-1}(\gamma(A)) + t)$. For general sets the same argument uses the lower one-sided derivative, or smooth approximation and passage to the limit; no differentiability of every parallel boundary is required.

If $\gamma(A) \geq 1/2$, then $\Phi^{-1}(\gamma(A)) \geq 0$. Finally,

$$1 - \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(t+u)^2/2}\, du \leq \frac{1}{2}e^{-t^2/2} \qquad (t \geq 0),$$

proving the asserted concentration estimate. □

Remark 6.68. *This set-concentration statement will imply Gaussian concentration for every Lipschitz function, including $x \mapsto |x|$.*

Remark 6.69. *The enlargement and perimeter formulations are equivalent: differentiating the enlargement inequality at $t = 0$ gives the perimeter bound, while the preceding argument proves the converse.*

Remark 6.70. *This means that if $Z \sim N(0, I)$, then*

$$\mathbb{P}(Z \in A) \geq 1/2 \Rightarrow \mathbb{P}\left(Z \notin A_t\right) \leq 1/2e^{-t^2/2}$$

Corollary 6.71. *Consider a Borel measurable set $A \subseteq \mathbb{S}^{n-1}$ and $A_t = \{x \in S^{n-1} : \operatorname{dist}_{\mathbb{S}^{n-1}}(x, A) \leq t\}$. Suppose $\mathbb{P}(\theta \in A) \geq 0.51$, where $\theta \sim \operatorname{Unif}(\mathbb{S}^{n-1})$. Then, $\mathbb{P}\left(\theta \notin A_t\right) \leq 2e^{-ct^2 n}$, where $c > 0$ is an absolute constant.*

Remark 6.72. *We define the spherical distance between two points $\theta, \eta \in \mathbb{S}^{n-1}$ as*

$$\operatorname{dist}_{\mathbb{S}^{n-1}}(\theta, \eta) := \sphericalangle(\theta, \eta).$$

The spherical distance between two points of $\mathbb{S}^{n-1}$ is at most π.

Proof sketch of Corollary 6.71. Write $G = R\theta$, where $G \sim \mathcal{N}(0, \mathrm{Id})$, $R = |G|$, and θ is uniform on $\mathbb{S}^{n-1}$, independently of R. Choose a small absolute constant $a > 0$ such that $\mathbb{P}(R \geq a\sqrt{n}) \geq 0.99$. For instance, the elementary Gaussian small-ball bound $\mathbb{P}(R \leq a\sqrt{n}) \leq (ea^2)^{n/2}$ permits such a choice. Set

$$B = \{r\theta : r \geq a\sqrt{n},\ \theta \in A\}.$$

Independence gives $\gamma(B) \geq 0.99 \cdot 0.51 > 1/2$.

If $|x - y| \leq s$ with $y \in B$ and $s < a\sqrt{n}/2$, then

$$\left|\frac{x}{|x|} - \frac{y}{|y|}\right| \leq \frac{2|x-y|}{|y|}, \qquad \operatorname{dist}_{\mathbb{S}^{n-1}}\left(\frac{x}{|x|}, \frac{y}{|y|}\right) \leq \frac{\pi s}{a\sqrt{n}}.$$

Thus, for $0 < t \leq 1$ and $s = a\sqrt{n}\, t/\pi$, the Euclidean enlargement B_s is contained in the cone over the spherical enlargement A_t. Consequently,

$$\sigma(A_t) \geq \gamma(B_s) \geq 1 - \tfrac{1}{2}e^{-s^2/2} \geq 1 - 2e^{-cnt^2}.$$

For $1 < t \leq \pi$, use the estimate at $t = 1$ and decrease c by an absolute factor. For $t > \pi$ the complement is empty. This proves the claim. □

Homework: add the remaining details to the proof and remove the $\sigma(A) \geq 0.51$ assumption, particularly, replace it with $\sigma(A) \geq 0.5$.

Definition 6.73. Let μ be a probability measure on a metric space (X, d). Then we define $\alpha_\mu(t)$ as:

$$\alpha_\mu(t) := \sup_{\substack{A \subseteq X \text{ measurable} \\ \mu(A) \geq 1/2}} \big(1 - \mu(A_t)\big),$$

where, as usual, $A_t = \{x \in X : d(x, A) \leq t\}$.

Gaussian isoperimetry gives the exact formula $\alpha_\gamma(t) = 1 - \Phi(t) \leq \frac{1}{2}e^{-t^2/2}$. After the improvement from 0.51 to 1/2 in the preceding exercise, spherical concentration gives $\alpha_\sigma(t) \leq 2e^{-cnt^2}$, where σ is uniform measure on $\mathbb{S}^{n-1}$.

Theorem 6.74. *Let (X, d, μ) be a probability metric space, and $f\colon X \to \mathbb{R}$ a random variable which is 1-Lipschitz (i.e. $\forall x, y \in X$ we have $|f(x) - f(y)| \leq d(x, y)$). Then,*

$$\mu(x \in X : |f(x) - \mathrm{med}(f)| > t) \leq 2\alpha_\mu(t).$$

Recall that $\mathrm{med}(f)$ is **the median** of f, that is, a number in $\mathbb{R}$ such that $\mu(f \geq \mathrm{med}(f)) \geq 1/2$ and $\mu(f \leq \mathrm{med}(f)) \geq 1/2$. What this theorem states basically is: "Lipschitz functions are almost constant on metric-measure spaces whose concentration function decays fast".

Remark 6.75. *Conversely, distance functions recover set concentration. If $\mu(A) \geq 1/2$, the function $f(x) = d(x, A)$ is 1-Lipschitz and has 0 as a median. Thus*

$$\alpha_\mu(t) \leq \sup_{\substack{\mathrm{Lip}(f) \leq 1 \\ m \text{ a median of } f}} \mu\{x : |f(x) - m| > t\}.$$

Proof of Theorem 6.74. Put $m = \mathrm{med}(f)$ and consider $A = \{f \geq m\}$, $B = \{f \leq m\}$. Both sets have measure at least 1/2. If $y \in A_t$, the Lipschitz property gives $f(y) \geq m - t$; if the distance infimum is not attained, use points whose distances decrease to it. Similarly, $y \in B_t$ implies $f(y) \leq m + t$. Hence

$$\{y : |f(y) - m| > t\} \subseteq A_t^c \cup B_t^c.$$

The union bound and the definition of α_μ give

$$\mu\{|f - m| > t\} \leq \mu(A_t^c) + \mu(B_t^c) \leq 2\alpha_\mu(t).$$

□

Homework: Prove that we can replace the median by the mean when μ is the Gaussian measure or the uniform measure on $\mathbb{S}^{n-1}$:

- $\gamma(x \in \mathbb{R}^n : |f(x) - \mathbb{E}_\gamma(f)| \geq t) \leq 2e^{-ct^2}$,
- $\sigma(x \in \mathbb{S}^{n-1} : |f(x) - \mathbb{E}_\sigma f| \geq t) \leq 2e^{-ct^2 n}$,

for any 1-Lipschitz function f.

Application 1. The Johnson-Lindenstrauss Lemma. Consider a set of points in $\mathbb{R}^n$, with n large, we can think of these points as data. We would like to reduce the complexity of the data by "reducing the dimension" of the data projecting it to a subspace, but not losing too much information about the geometry of the point set.

Theorem 6.76 (Johnson-Lindenstrauss Lemma)**.** *Let $\mathcal{X}$ be a set of N points in $\mathbb{R}^n$. Take $0 < \epsilon < 1$, and let $m \leq n$ be an integer satisfying $m \geq C\epsilon^{-2} \log N$, for a sufficiently large absolute constant C. Now, let E be a random subspace of $\mathbb{R}^n$ of dimension m sampled uniformly from $G_{n,m}$, the collection of all m-dimensional subspaces of $\mathbb{R}^n$. Let also P_E be the orthogonal projection on E, and $Q = \sqrt{\frac{n}{m}} P_E$ the normalized projection. Then, with probability at least $1 - 2e^{-c\epsilon^2 m}$, we have that for all $x, y \in \mathcal{X}$:*

$$(1 - \epsilon) |x - y| \leq |Qx - Qy| \leq (1 + \epsilon) |x - y| . \tag{70}$$

Rephrased in a general way the theorem states that "with high probability the normalized random projection is almost an isometry".

Corollary 6.77. *There exists a subspace of dimension m whose normalized orthogonal projection satisfies (70).*

Remark 6.78. *We call the collection of all m-dimensional subspaces of $\mathbb{R}^n$ as $G_{n,m}$. The random subspace is selected uniformly in the previous theorem.*

Lemma 6.79. *Let P be the orthogonal projection onto a uniform random m-dimensional subspace of $\mathbb{R}^n$, where $1 \leq m \leq n$. Fix $z \in \mathbb{R}^n$ and $0 < \epsilon < 1$. Then*

1. $\mathbb{E} |Pz|^2 = (m/n)|z|^2$;
2. *with probability at least $1 - 2e^{-c\epsilon^2 m}$,*

$$(1 - \epsilon)\sqrt{\frac{m}{n}}|z| \leq |Pz| \leq (1 + \epsilon)\sqrt{\frac{m}{n}}|z|.$$

Proof. By homogeneity, assume $|z| = 1$. Rotation invariance lets us fix $F = \operatorname{span}(e_1, \ldots, e_m)$ and replace z by $Z \sim \operatorname{Unif}(\mathbb{S}^{n-1})$. Thus

$$\mathbb{E} |Pz|^2 = \mathbb{E} |P_F Z|^2 = \sum_{i=1}^{m} \mathbb{E}\, Z_i^2 = \frac{m}{n}, \tag{71}$$

since $\sum_{i=1}^{n} Z_i^2 = 1$ and all coordinates have the same distribution.

For the second assertion, $g(Z) = |P_F Z|$ is 1-Lipschitz. The spherical concentration estimate around the mean gives

$$\mathbb{P}(|g - \mathbb{E}\, g| > u) \leq 2e^{-cnu^2}, \qquad \operatorname{Var}(g) \leq \frac{C}{n}.$$

Since $\mathbb{E}\, g^2 = m/n$,

$$0 \le \sqrt{m/n} - \mathbb{E}\, g = \frac{\mathrm{Var}(g)}{\sqrt{m/n} + \mathbb{E}\, g} \le \frac{C}{\sqrt{mn}}.$$

When $\epsilon^2 m$ exceeds a sufficiently large absolute constant,

$$\sqrt{m/n} - \mathbb{E}\, g \le (\epsilon/2)\sqrt{m/n}.$$

Apply concentration with $u = (\epsilon/2)\sqrt{m/n}$ to get the result. When $\epsilon^2 m$ is smaller, choose the absolute constant c so that $2e^{-c\epsilon^2 m} \ge 1$; the asserted probability bound is then automatic. Scaling restores the factor $|z|$. □

Proof of the Johnson–Lindenstrauss Lemma. For each nonzero $z = x - y$ with $x, y \in \mathcal{X}$, let

$$E_z = \{(1-\epsilon)|z| \le |Qz| \le (1+\epsilon)|z|\}.$$

The preceding lemma gives $\mathbb{P}(E_z^c) \le 2e^{-c\epsilon^2 m}$. There are at most N^2 difference vectors, so

$$\mathbb{P}\left(\bigcap_z E_z\right) \ge 1 - 2N^2 e^{-c\epsilon^2 m} \ge 1 - 2e^{-c'\epsilon^2 m},$$

provided $m \ge C\epsilon^{-2}\log N$ with C sufficiently large. The zero difference causes no restriction. □

Application 2. Milman-Dvoretzky Theorem (1971).

Theorem 6.80 (Milman–Dvoretzky theorem, with the Gordon–Schechtman dependence on ϵ)**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and $0 < \epsilon < 1$. There exist an invertible linear map T, a scale $\mu > 0$, and an integer $k \ge c\epsilon^2 \log n$ such that, for a uniform random k-dimensional subspace F, with probability at least $1 - e^{-c_1\epsilon^2 k}$,*

$$(1-\epsilon)\mu|x| \le \|Tx\| \le (1+\epsilon)\mu|x| \qquad \text{for every } x \in F.$$

Here $c, c_1 > 0$ are absolute constants, and the choice of T is made before choosing F.

After a suitable fixed change of coordinates, a random subspace of the indicated dimension is almost Euclidean. The stated dependence on ϵ uses the refinement of Gordon and Schechtman [12; 42]. The elementary net argument below explains the mechanism and proves a version with an additional factor $1/\log(C/\epsilon)$ in the dimension; it is not a complete proof of the sharper dependence stated above.

The elementary argument is based on concentration for Lipschitz functions and a net argument. We first prove the corresponding intermediate lemmas and propositions. Throughout the following discussion we denote $b > 0$ as smallest positive number such that $\|x\| \le b|x|$ for all $x \in \mathbb{R}^n$. Let $M := \int_{\mathbb{S}^{n-1}} \|x\|\, d\sigma(x)$ where σ is the uniform measure on $\mathbb{S}^{n-1}$. Let $O(n)$ be the n-dimensional orthogonal group and $\mathcal{V}_n$ be the uniform measure on $O(n)$. A useful fact is that the probability of hitting the area A on the sphere is equal to the probability that any point on the sphere is randomly rotated and lands into the area A. That is, for any fixed $x_0 \in \mathbb{S}^{n-1}$ and $A \subset \mathbb{S}^{n-1}$, we have $\sigma(A) = \mathcal{V}_n(\{u \in O(n) : ux_0 \in A\})$.

To begin with, we are going to prove the following lemma.

Lemma 6.81. *Fix $m \in \mathbb{N}$ such that $m \leq \frac{1}{4}e^{c_1\epsilon^2 n/2}$. Let $\{y_1, ..., y_m\}$ be a collection of m points in $\mathbb{S}^{n-1}$. Then there exists $B \subset O(n)$ with $\mathcal{V}_n(B) \geq 1 - \exp\{-c_1\epsilon^2 n/2\}$, such that for any $u \in B$ and any $i \in \{1, ..., m\}$, we have*

$$M - b\epsilon \leq \|uy_i\| \leq M + b\epsilon. \tag{72}$$

Proof of Lemma 6.81. Let $A = \{x \in \mathbb{S}^{n-1} : M - b\epsilon \leq \|x\| \leq M + b\epsilon\}$ and let $f(x) = \|x\|$. Note that $f(x)$ is b-Lipschitz and $\mathbb{E}\, f(X) = M$. That means we can apply concentration of Lipschitz functions around their mean, following Theorem 6.74 and the subsequent exercise, and get

$$\sigma(A) \geq 1 - 4\exp\{-c_1\epsilon^2 n\}. \tag{73}$$

Now consider $B_i = \{u \in O(n) : M - b\epsilon \leq \|uy_i\| \leq M + b\epsilon\}$ and define our B to be $B = \cap_{i=1}^m B_i$. Recall that we have $\mathcal{V}_n(B_i) = \sigma(A)$ from previous argument. By union bound we conclude

$$\mathcal{V}_n(B) = 1 - \mathcal{V}_n(B^c) \geq 1 - \sum_{i=1}^m \mathcal{V}_n(B_i^c) \geq 1 - 4m\exp\{-c_1\epsilon^2 n\} \geq 1 - \exp\{-c_1\epsilon^2 n/2\}, \tag{74}$$

where in the last step we recall $m \leq 1/4\exp\{c_1\epsilon^2 n/2\}$. □

The lemma above basically states the concentration of the norm of randomly rotated vectors on a sphere. It can be used to obtain similar results for vectors in subspaces, shown in the following proposition.

Proposition 6.82. *Fix $\delta, \epsilon \in (0,1), k \in \mathbb{N}$ and suppose $(1 + \frac{2}{\delta})^k \leq \frac{1}{4}\exp\{c_1\epsilon^2 n/2\}$. Then there exists a set of k-dimensional subspaces, call it Γ, with $\mathcal{V}_{n,k}(\Gamma) \geq 1 - \exp\{-c_1\epsilon^2 n/2\}$[1], such that for every subspace $F \in \Gamma$, there exists a δ-net $\mathcal{N}_F$ on $\mathbb{S}^{n-1} \cap F$, so that*

$$\forall x \in \mathcal{N}_F, \quad M - b\epsilon \leq \|x\| \leq M + b\epsilon. \tag{75}$$

Proof of Proposition 6.82. Fix a k-dimensional subspace F_0. Consider the δ-net $\{y_1, ..., y_m\}$ on $\mathbb{S}^{n-1} \cap F_0$. Since $\mathbb{S}^{n-1} \cap F_0$ has dimension $k-1$, the net argument gives $m \leq (1 + \frac{2}{\delta})^k$. Now applying Lemma 6.81 to $\{y_1, ..., y_m\}$, we can find $B \subset O(n)$ such that

$$\forall u \in B, \forall i \in \{1, ..., m\}, \quad M - b\epsilon \leq \|uy_i\| \leq M + b\epsilon. \tag{76}$$

For a rotation $u \in B$, define $x_i = uy_i$ and $F_u = uF_0$. Note that $\{x_i\}$ forms a δ-net for $\mathbb{S}^{n-1} \cap F_u$, since orthogonal transformation preserves Euclidean norm.

Now construct Γ by $\Gamma = \{F_u : u \in B\}$. We have $\mathcal{V}_{n,k}(\Gamma) \geq \mathcal{V}_n(B) \geq 1 - \exp\{-c_1\epsilon^2 n/2\}$. For every $F_u \in \Gamma$, the δ-net is exactly $\{x_1, ..., x_m\} = \{uy_1, ..., uy_m\}$ and it satisfies

$$M - b\epsilon \leq \|x_i\| = \|uy_i\| \leq M + b\epsilon \tag{77}$$

by construction. □

[1] $\mathcal{V}_{n,k}$ is the uniform measure on all k-dimensional linear subspaces. In other words, here it is equivalent to say "we take a random subspace F, with probability at least $1 - \exp\{-c_1\epsilon^2 n/2\}$, ..."

The Proposition above is close to what we want for concentration of norm, except that it only works for δ-net. Now let us extend it to any points in the subspace.

Proposition 6.83. *Fix $\delta, \epsilon \in (0,1), k \in \mathbb{N}$ and suppose $(1+\frac{2}{\delta})^k \le \frac{1}{4}\exp\{c_1\epsilon^2 n/2\}$. If we take a random k-dimensional subspace F, then with probability at least $1-\exp\{-c_1\epsilon^2 n/2\}$, for all $y \in F \cap \mathbb{S}^{n-1}$,*

$$\frac{1-2\delta}{1-\delta}M - \frac{b\epsilon}{1-\delta} \le \|y\| \le \frac{M+b\epsilon}{1-\delta}. \tag{78}$$

Proof of Proposition 6.83. Let F have the net supplied by Proposition 6.82, and put $B_F = \max_{y\in F\cap\mathbb{S}^{n-1}} \|y\|$. For each such y, choose $x \in \mathcal{N}_F$ with $|y-x| \le \delta$. Since $y - x \in F$,

$$\|y\| \le \|x\| + \|y-x\| \le M + b\epsilon + \delta B_F.$$

Taking the maximum gives $B_F \le (M+b\epsilon)/(1-\delta)$, proving the upper bound. The reverse triangle inequality gives

$$\|y\| \ge \|x\| - \|y-x\| \ge M - b\epsilon - \delta\frac{M+b\epsilon}{1-\delta} = \frac{(1-2\delta)M - b\epsilon}{1-\delta},$$

which proves the lower bound. □

A further geometric input is that one can choose an invertible T so that, for the norm $x \mapsto \|Tx\|$,

$$\frac{M}{b} \ge c\sqrt{\frac{\log n}{n}}. \tag{79}$$

This position estimate is an additional ingredient, not proved by the preceding concentration argument. To see the dimension supplied by that argument, let η be the desired relative error and choose $\delta = \eta/8$ and $\epsilon = \eta M/(8b)$ in Proposition 6.83. The conclusion then lies between $(1-\eta)M$ and $(1+\eta)M$, while the net-size condition allows dimensions of order

$$\frac{\eta^2 n (M/b)^2}{\log(C/\eta)}.$$

Using (79), this gives the familiar elementary bound of order $\eta^2 \log n/\log(C/\eta)$, with integer rounding understood. Removing the logarithmic loss requires the sharper Gaussian comparison or chaining argument cited above; it does not follow just by optimizing the constants in this one-net proof.

Remark 6.84. *Equation (79) is sharp. For example $\|\cdot\| = \|\cdot\|_\infty$, in such case*

- *$M = \int_{\mathbb{S}^{n-1}} \max_i |x_i|\, d\sigma(x) \asymp \sqrt{\log n/n}$ for $n \ge 2$ (HW).*
- *$b = 1$ as $\|x\|_\infty \le |x|$.*

Remark 6.85. *The logarithmic order in n in Theorem 6.80, for fixed distortion, is sharp for $\|\cdot\|_\infty$. This does not assert sharpness of its dependence on ϵ.*

7 The mass transport method

Imagine that there are several bakeries producing croissants and shops waiting for the products. The productivity of bakeries and the demand of shops are different. What is the best way to deliver croissants, in terms of delivery distance? This can be formalized as the Monge-Kantorovich problem, or transportation problem.

7.1 Basic definitions and setup

Definition 7.1 (Transport map)**.** Let μ, ν be finite Borel measures on $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a Borel measurable map defined μ-almost everywhere. We say that T transports μ into ν if, for every Borel set $A \subset \mathbb{R}^n$,

$$\nu(A) = \mu(T^{-1}A), \tag{80}$$

where $T^{-1}A = \{x : T(x) \in A\}$ is the preimage of A. We write $T_\star \mu = \nu$; the notation $T_\# \mu = \nu$ has the same meaning.

Remark 7.2 (HW)**.** *Equivalently, for every $\phi \in L^1(\nu)$,*

$$\int \phi d\nu = \int \phi \circ T d\mu. \tag{81}$$

Example 7.3 (Discrete measure)**.** *Let $\mu = \sum_{i=1}^m \lambda_i \delta_{x_i}$ be a discrete measure supported on $\{x_1, ..., x_m\}$. For any map T, the pushforward is $\nu = \sum_{i=1}^m \lambda_i \delta_{T(x_i)}$.*

Example 7.4 (Gaussian measure)**.** *Let μ be standard Gaussian measure on $\mathbb{R}$ and ν uniform measure on $[0, 1]$. Let $T(x) = \Phi(x)$ be the Gaussian cdf. Then T transports μ into ν. Indeed, for any $0 < a < b < 1$, we have*

$$\begin{aligned} \nu([a, b]) &= b - a \\ \mu(T^{-1}([a, b])) &= \Phi(T^{-1}(b)) - \Phi(T^{-1}(a)) = b - a. \end{aligned} \tag{82}$$

7.2 Useful Properties of Lipschitz Transport Maps

Moving forward, we can extend the preceding example to $\mathbb{R}^n$.

Example 7.5 (Gaussian measure on $\mathbb{R}^n$)**.** *Let γ be the standard Gaussian measure in $\mathbb{R}^n$. Consider the transport map $T : \mathbb{R}^n \to \mathbb{R}^n$ given by*

$$T(x) = \begin{bmatrix} \Phi(x_1) \\ \vdots \\ \Phi(x_n) \end{bmatrix}, \tag{83}$$

where Φ is the cdf of the standard one-dimensional Gaussian distribution. We have $T_\star\gamma = Unif([0,1]^n)$, and T is $\frac{1}{\sqrt{2\pi}}$-Lipschitz continuous. Recall that a map $T : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous if there exists a real constant $L \geq 0$ such that

$$|T(x) - T(y)| \leq L|x - y|, \qquad \forall x, y \in \mathbb{R}^n.$$

Since $\Phi(s) = \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, we know

$$\Phi'(s) = \left(\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right)' = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \leq \frac{1}{\sqrt{2\pi}},$$

which implies T is $\frac{1}{\sqrt{2\pi}}$-Lipschitz continuous.

The following proposition demonstrates that an L-Lipschitz continuous transport map allows us to extend the properties of one measure to another.

Proposition 7.6. *Suppose μ, ν are probability measures on $\mathbb{R}^n$, T is L-Lipschitz for some $L > 0$, and $T_\star\mu = \nu$. Then the following hold:*

1. *$\alpha_\nu(Lt) \leq \alpha_\mu(t)$ for $t > 0$, where α is the concentration function.*

2. *The Poincaré inequality propagates:*

$$C_{poin}(\nu) \leq L^2 C_{poin}(\mu).$$

In other words, if every sufficiently regular f satisfies

$$\int f^2 d\mu - \left(\int f d\mu \right)^2 \leq C_{poin}(\mu) \int |\nabla f|^2 d\mu,$$

then every sufficiently regular g satisfies

$$\int g^2 d\nu - \left(\int g d\nu \right)^2 \leq L^2 C_{poin}(\mu) \int |\nabla g|^2 d\nu.$$

3. *The log-Sobolev inequality propagates. Recall the definition of entropy in Remark 6.49. If*

$$\mathrm{Ent}_\mu(f^2) \leq \frac{C}{2} \int |\nabla f|^2 d\mu,$$

then

$$\mathrm{Ent}_\nu(g^2) \leq \frac{L^2 C}{2} \int |\nabla g|^2 d\nu.$$

4. Isoperimetry propagates. Let

$$I_\mu(a) = \inf_{\substack{A\subset\mathbb{R}^n \text{ Borel} \\ \mu(A)=a}} \mu^+(\partial A), \qquad a \in [0,1],$$

where

$$\mu^+(\partial A) = \liminf_{t\downarrow 0} \frac{\mu(A_t \setminus A)}{t}, \qquad A_t = \{x : \operatorname{dist}(x, A) \le t\}.$$

Then $I_\nu(a) \ge L^{-1} I_\mu(a)$.

Proof sketch. The useful observation is the preimage inclusion

$$(T^{-1}B)_t \subset T^{-1}(B_{Lt}), \qquad t > 0. \tag{84}$$

Indeed, points within distance t of $T^{-1}B$ are sent within distance Lt of B. This does not require T to be invertible.

1. If $\nu(B) \ge 1/2$, then $\mu(T^{-1}B) \ge 1/2$, so

   $$\nu(B_{Lt}) \ge \mu((T^{-1}B)_t) \ge 1 - \alpha_\mu(t).$$

   Taking the infimum over B proves the concentration comparison.

2. Apply the Poincaré inequality for μ to $f = g \circ T$. The pushforward identity gives

   $$\operatorname{Var}_\nu(g) = \operatorname{Var}_\mu(g \circ T),$$

   while the chain rule gives

   $$\int |\nabla(g\circ T)|^2 d\mu \le L^2 \int |\nabla g(Tx)|^2 d\mu(x) = L^2 \int |\nabla g|^2 d\nu. \tag{85}$$

3. Use (85) and $\operatorname{Ent}_\nu(g^2) = \operatorname{Ent}_\mu((g\circ T)^2)$.

4. The proof is left as HW: apply (84), subtract $\mu(T^{-1}B) = \nu(B)$, divide by t, and take the lower limit.

For presentation, the gradient calculation is written for smooth maps and functions. The Lipschitz version follows using local Lipschitz constants (and approximation when ordinary gradients are used). □

The following corollary demonstrates the good properties of uniform measures on the cube.

Corollary 7.7. *Let* $d\nu = \mathbb{1}_{[0,1]^n}(x)dx$. *Then we have*

1. $\alpha_\nu(t) \le 1 - \Phi(\sqrt{2\pi}\, t) \le \frac{1}{2} e^{-\pi t^2}$ *for all* $t > 0$.

2. *For every smooth $f : [0,1]^n \to \mathbb{R}$, we have*

$$\int_{[0,1]^n} f^2 dx - \left(\int_{[0,1]^n} f dx \right)^2 \leq \frac{1}{2\pi} \int_{[0,1]^n} |\nabla f|^2 dx. \tag{86}$$

This transport bound is not sharp: the optimal constant is $1/\pi^2$, attained, for example, by $f(x) = \cos(\pi x_1)$.

3. *For every sufficiently regular f, the Gaussian log-Sobolev inequality gives $\mathrm{Ent}_\nu(f^2) \leq \frac{1}{\pi} \int |\nabla f|^2 d\nu$.*

4. *For any $A \subset \mathbb{R}^n$ with $|A \cap [0,1]^n| = a \in [0,1]$, we have*

$$\nu^+(\partial A) \geq \sqrt{2\pi} I_\gamma(a) = e^{-\frac{(\Phi^{-1}(a))^2}{2}}, \tag{87}$$

where $I_\gamma(a) = (2\pi)^{-1/2} e^{-(\Phi^{-1}(a))^2/2}$ for $a \in (0,1)$, with the endpoint values understood by continuity. For a halfspace whose boundary cuts through the cube, this perimeter is the area of the corresponding section.

Remark 7.8. *The constant in statement (4) is sharp at volume $1/2$: take $A = \{x : x_1 \leq 1/2\}$.*

Next, we present a special example that was previously studied by Vaaler, Hadwiger, and others.

Corollary 7.9 (The hyperplane case of Vaaler's theorem on central sections of the cube)**.** *For any $\theta \in \mathbb{S}^{n-1}$, we have $|[-1/2, 1/2]^n \cap \theta^\perp|_{n-1} \geq 1$.*

Indeed, a hyperplane through the center bisects the cube. Apply the preceding perimeter bound at $a = 1/2$. At this point, we can conclude that it is nice to have a Lipschitz map transporting nice measures into measures. Now the question is, are there other measures besides the uniform measure of the cube where we can construct a Lipschitz map transporting between measures?

7.3 Optimal Transport Map with respect to quadratic cost

Definition 7.10 (Optimal Transport)**.** A map T_0 transporting μ into ν is said to be **optimal w.r.t. quadratic cost** if for all T with $T_\star \mu = \nu$,

$$\int_{\mathbb{R}^n} |T_0 x - x|^2 d\mu \leq \int_{\mathbb{R}^n} |Tx - x|^2 d\mu. \tag{88}$$

In other words, T_0 minimizes the average squared distance that the mass travels.

Now that we have defined the optimal transport map (w.r.t. quadratic cost), the question is, how can we check if a map attains optimality? Also, how can we explicitly construct an optimal transport map between measures? To answer these questions, we first state an important theorem in optimal transport.

Theorem 7.11 (Brenier's theorem)**.** *Let μ, ν be Borel probability measures with finite second moments, and suppose that μ is absolutely continuous with respect to Lebesgue measure. A map T with $T_\star \mu = \nu$ is optimal for quadratic cost if and only if*

$$T = \nabla F \quad \mu\text{-almost everywhere}$$

for a proper lower semicontinuous convex function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, finite μ-almost everywhere. Such a map exists and is unique up to μ-null sets. It is called the Brenier map*.*

In the following, we give some examples of the construction of optimal transport maps.

Example 7.12. *Let γ be the standard Gaussian measure on $\mathbb{R}^n$. Suppose we have*

$$T_\star \gamma = \frac{1}{2}\delta_{e_1} + \frac{1}{2}\delta_{-e_1},$$

where e_1 is the first coordinate vector and δ_x is the Dirac measure at $x \in \mathbb{R}^n$. The optimal transport map w.r.t. quadratic cost, in this case, is given by

$$T(x) = \operatorname{sign}(x_1) e_1. \tag{89}$$

We can see that $T(x) = \nabla F(x)$ where $F(x) = |x_1|$ and $\nabla F(x) = \operatorname{sign}(x_1) e_1$.

Example 7.13. *Let us consider a more general example. Suppose μ is an absolutely continuous probability measure with finite second moment, and the target is*

$$T_\star \mu = \sum_{i=1}^{N} \lambda_i \delta_{v_i}, \qquad v_i \in \mathbb{S}^{n-1} \text{ distinct}, \quad \lambda_i > 0, \quad \sum_i \lambda_i = 1.$$

The optimal map has the form $T = \nabla F$, where

$$F(x) = \max_{1 \le j \le N} \big(y_j + \langle x, v_j \rangle \big).$$

Here the scalar constants y_j are chosen so that the region where the jth affine function is maximal has μ-mass λ_j. The gradients are the target points v_j; the weights determine the sizes of these regions, not the slopes.

Remark 7.14 (Caffarelli's theorem)**.** *Let γ be the standard Gaussian measure, and let $d\mu = e^{-V(x)} dx$ be a probability measure with $V \in C^2(\mathbb{R}^n)$ and $\nabla^2 V \ge k I_n$ for some $k > 0$. The Brenier map $T_\star \gamma = \mu$ is $1/\sqrt{k}$-Lipschitz. Such measures are called* uniformly log-concave*. We will prove Caffarelli's theorem later.*

Before we prove Brenier's theorem, let us first sketch the proof for the discrete version of Brenier.

Theorem 7.15 (Discrete version of Brenier's theorem)**.** *Let $x_1, \dots, x_N, y_1, \dots, y_N \in \mathbb{R}^n$.*

1. *Suppose there exists a convex function $F : \mathbb{R}^n \to \mathbb{R}$, differentiable at $x_1, \ldots, x_N$, such that*

$$\nabla F(x_i) = y_i, \qquad \forall i \in \{1, \ldots, N\}. \tag{90}$$

Then for any permutation σ on the set $\{1, \ldots, N\}$, we have

$$\sum_{i=1}^{N} |x_i - y_{\sigma(i)}|^2 \geq \sum_{i=1}^{N} |x_i - y_i|^2. \tag{91}$$

2. *Vice versa, if*

$$\sum_{i=1}^{N} |x_i - y_i|^2 = \min_{\sigma \in S_N} \sum_{i=1}^{N} |x_i - y_{\sigma(i)}|^2, \tag{92}$$

then there exists a convex function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that

$$y_i \in \partial F(x_i), \qquad \forall i \in \{1, \ldots, N\}. \tag{93}$$

Here, we would like to introduce the concept of sub-gradient of a convex function:

Definition 7.16 (Sub-gradient of a convex function f)**.** The subdifferential of a convex function f at a point x where $f(x) < \infty$ is the set of its subgradients:

$$\partial f(x) = \{y \in \mathbb{R}^n : f(z) \geq f(x) + \langle y, z - x \rangle, \quad \forall z \in \mathbb{R}^n\}. \tag{94}$$

Note that if $f \in C^1(\mathbb{R}^n)$, then $\partial f(x) = \{\nabla f(x)\}$. Let us consider a quick example.

Example 7.17 (Sub-gradient of the absolute value function)**.** *Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(t) = |t|$. The sub-gradient of f at 0 is $\partial f(0) = [-1, 1]$. The sub-gradient of f on points not equal to 0 is*

$$\partial f(x) = \begin{cases} \{-1\}, & x < 0, \\ \{1\}, & x > 0. \end{cases}$$

Now that we have introduced the concept of sub-gradient, we present the proof of Theorem 7.15.

Proof of Discrete version of Brenier's theorem (Theorem 7.15). Let us first prove statement (1) in the theorem.

Proof of Statement (1)

Given that F is convex, at every differentiability point x we know

$$F(y) - F(x) \geq \langle \nabla F(x), y - x \rangle, \qquad y \in \mathbb{R}^n, \tag{95}$$

Also, since $\nabla F(x_i) = y_i$ for $i \in \{1, \dots, N\}$, it follows from (95) and the commutativity of addition that

$$\sum_{i=1}^{N} \langle \nabla F(x_i), x_{\sigma(i)} - x_i \rangle \leq \sum_{i=1}^{N} \big(F(x_{\sigma(i)}) - F(x_i) \big) = 0$$

This implies that

$$\sum_{i=1}^{N} \langle \nabla F(x_i), x_{\sigma(i)} \rangle \leq \sum_{i=1}^{N} \langle \nabla F(x_i), x_i \rangle, \tag{96}$$

or equivalently, $\sum_{i=1}^{N} \langle y_i, x_{\sigma(i)} \rangle \leq \sum_{i=1}^{N} \langle y_i, x_i \rangle$. This further implies that

$$\sum_{i=1}^{N} |y_i - x_{\sigma(i)}|^2 \geq \sum_{i=1}^{N} |y_i - x_i|^2 \tag{97}$$

because $\sum_{i=1}^{N} |x_{\sigma(i)}|^2 = \sum_{i=1}^{N} |x_i|^2$. Replacing σ by its inverse gives the stated form of (1). We will obtain (2) below from cyclic monotonicity and Rockafellar's theorem.

□

Definition 7.18 (Coupling of measures)**.** Let μ, ν be probability measures on $\mathbb{R}^n$. A coupling of μ and ν is a Borel probability measure γ on $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are μ and ν. Equivalently, for every bounded Borel measurable function φ,

$$\int \varphi(x) d\gamma(x, y) = \int \varphi(x) d\mu(x), \qquad \int \varphi(y) d\gamma(x, y) = \int \varphi(y) d\nu(y).$$

Example 7.19. *If $T_\star \mu = \nu$, the associated coupling is*

$$\gamma_T = (\mathrm{Id}, T)_\star \mu.$$

Thus γ_T is concentrated on the graph $\{(x, T(x))\}$: it describes moving all the mass at x to the single point $T(x)$. A general coupling may split the mass at a point among several destinations.

7.4 Monge Problem

Definition 7.20. (Monge Problem) Let a cost function $C : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ be lower semicontinuous. For probability measures μ, ν, the Monge problem is

$$\inf_{T_* \mu = \nu} \int_{\mathbb{R}^n} C(x, Tx) d\mu(x) = C_\mu(\mu, \nu) \tag{98}$$

If we use a quadratic cost, i.e.

$$C(x, y) = |x - y|^2 \tag{99}$$

then a map attaining the infimum in (98) is optimal for quadratic cost. If no transport map exists, the infimum over maps is $+\infty$.

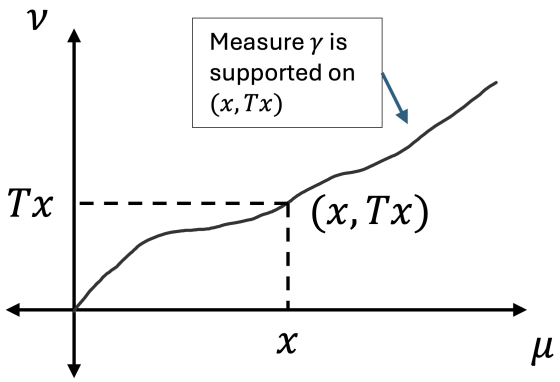


Figure 3: A coupling associated with a transport map.

7.5 Kantorovich Problem

Let μ and ν be probability measures on $\mathbb{R}^n$. The Kantorovich problem is

$$C_k(\mu,\nu) = \inf_{\gamma} \int_{\mathbb{R}^n \times \mathbb{R}^n} C(x,y) d\gamma(x,y) \tag{100}$$

where γ is a coupling of μ and ν, defined on $\mathbb{R}^n \times \mathbb{R}^n$. Every transport map induces a coupling, so $C_k(\mu,\nu) \leq C_\mu(\mu,\nu)$. Equality holds, in particular, for quadratic cost when μ is absolutely continuous and both measures have finite second moments, by Brenier's theorem.

Theorem 7.21. *(Without proof) Let a cost function $C : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ be lower semi-continuous. For any Borel probability measures μ, ν, there exists a coupling γ^* attaining the Kantorovich infimum. Its cost is finite if at least one coupling has finite cost.*

7.6 Cyclic Monotonicity

Definition 7.22. A set $\Lambda \subset \mathbb{R}^n \times \mathbb{R}^n$ is C-cyclically monotone if, for every finite sequence $(x_i, y_i)_{i=1}^N$ in Λ, with $x_{N+1} = x_1$,

$$\sum_{i=1}^{N} C(x_i, y_i) \leq \sum_{i=1}^{N} C(x_{i+1}, y_i).$$

Theorem 7.23. *If γ^* is an optimal coupling of finite cost for a continuous nonnegative cost $C(x,y)$, then* $\mathrm{supp}(\gamma^*)$ *is C-cyclically monotone.*

Recall that the support of a measure consists of the points every open neighborhood of which has positive measure.

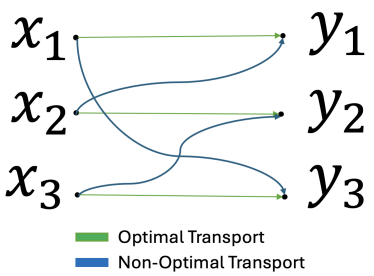


Figure 4: Cyclic monotonicity: cycling the destinations cannot lower the cost.

Proof. Homework: if a cycle strictly lowers the cost, continuity gives small neighborhoods on which the same improvement holds. Reassign a sufficiently small equal amount of mass around these neighborhoods to contradict optimality. □

Remark 7.24. *For quadratic cost, expanding the squares and using $x_{N+1} = x_1$ gives*

$$\sum_{i=1}^{N} |x_i - y_i|^2 \le \sum_{i=1}^{N} |x_{i+1} - y_i|^2 \quad \Longleftrightarrow \quad \sum_{i=1}^{N} \langle x_i, y_i \rangle \ge \sum_{i=1}^{N} \langle x_{i+1}, y_i \rangle$$
$$\Longleftrightarrow \quad \sum_{i=1}^{N} \langle x_i - x_{i+1}, y_i \rangle \ge 0.$$

We call this simply cyclic monotonicity.

Our aim is to relate cyclic monotonicity to convex functions. In dimension one, it is equivalent to monotonicity of the relation:

$$(x - x')(y - y') \ge 0 \qquad \text{for all } (x, y), (x', y') \in S.$$

The relation may be multivalued at a point, as happens for the subdifferential of $|x|$ at zero. A nondecreasing function is the derivative almost everywhere of a convex function; differentiability everywhere is not required. Checking these assertions is a useful exercise.

Recall that the subdifferential of a convex function φ is

$$\partial\varphi(x) = \{y : \varphi(z) \ge \varphi(x) + \langle y, z - x \rangle \text{ for every } z\}, \qquad \varphi(x) < \infty.$$

We also write

$$\partial\varphi = \{(x, y) : y \in \partial\varphi(x)\} \subset \mathbb{R}^n \times \mathbb{R}^n$$

for its graph. At a differentiability point, $\partial\varphi(x) = \{\nabla\varphi(x)\}$.

Theorem 7.25 (Rockafellar's theorem)**.** *A nonempty set $S \subset \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone if and only if $S \subset \partial\varphi$ for some proper lower semicontinuous convex function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.*

Proof. Subgradients of convex functions are cyclically monotone. Indeed, if $y_i \in \partial\varphi(x_i)$, then

$$\varphi(x_{i+1}) \geq \varphi(x_i) + \langle y_i, x_{i+1} - x_i \rangle.$$

Summing in i cancels the function values and gives $\sum_i \langle y_i, x_{i+1} - x_i \rangle \leq 0$.

Conversely, fix a base point $(x_0, y_0) \in S$ and define

$$\varphi(x) = \sup_{\substack{m \geq 0 \\ (x_i, y_i) \in S,\ 1 \leq i \leq m}} \left\{ \sum_{i=0}^{m-1} \langle y_i, x_{i+1} - x_i \rangle + \langle y_m, x - x_m \rangle \right\}.$$

The term with $m = 0$ is $\langle y_0, x - x_0 \rangle$. Here are the important points:

- φ is convex and lower semicontinuous because it is a supremum of affine functions.
- Cyclic monotonicity gives $\varphi(x_0) \leq 0$, while the term $m = 0$ gives $\varphi(x_0) \geq 0$. Thus $\varphi(x_0) = 0$, so φ is proper.
- If $(a, b) \in S$, closing a chain by (a, b) and then returning to x_0 gives $\varphi(a) \leq \langle b, a - x_0 \rangle < \infty$. Appending (a, b) to a chain ending at a gives $\varphi(z) \geq \varphi(a) + \langle b, z - a \rangle$. Hence $b \in \partial\varphi(a)$.

Notice that φ need not be finite everywhere, but it is finite at every first coordinate of a point in S. □

This also proves statement (2) of Theorem 7.15: optimality under all permutations implies cyclic monotonicity of the finite set $\{(x_i, y_i)\}$, and Rockafellar's theorem then supplies the convex function.

7.7 Kantorovich Duality

Definition 7.26. (Legendre Transform) For a proper convex function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, its Legendre transform is

$$\varphi^*(x) = \sup_{y \in \mathbb{R}^n} \left(\langle x, y \rangle - \varphi(y) \right) \tag{101}$$

7.7.1 Examples

Example 7.27.

$$f(x) = \frac{x^2}{2}$$

for $x \in \mathbb{R}$. Its transform is

$$f^*(x) = \left(\frac{x^2}{2} \right)^* = \frac{x^2}{2}$$

Example 7.28.

$$\varphi(x) = |x|$$

for $x \in \mathbb{R}^n$. Its transform is

$$\varphi^*(x) = \begin{cases} 0 & ; |x| \le 1 \\ \infty & ; |x| > 1 \end{cases}$$

Example 7.29.

$$\left(\frac{|t|^p}{p}\right)^* = \frac{|t|^q}{q}$$

where $1 < p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Notes:

- If φ is proper, lower semicontinuous, and convex, then $\varphi^{**} = \varphi$.
- For all $x, y \in \mathbb{R}^n$, Fenchel's inequality says

$$\varphi(x) + \varphi^*(y) \ge \langle x, y \rangle. \tag{102}$$

  Equality holds exactly when $y \in \partial\varphi(x)$.

- In particular, if φ is differentiable at x, then the supremum defining $\varphi^*(\nabla\varphi(x))$ is attained at x, and

$$\varphi(x) + \varphi^*(\nabla\varphi(x)) = \langle x, \nabla\varphi(x) \rangle.$$

  The proof is left as HW: rearrange the defining subgradient inequality.

Theorem 7.30 (Kantorovich duality)**.** *Let μ, ν be probability measures on $\mathbb{R}^n$ with finite second moments. Then*

$$\begin{aligned} \min_{\gamma} \int -\langle x, y \rangle d\gamma(x, y) &= \max_{\varphi, \psi} \left(-\int \varphi \, d\mu - \int \psi \, d\nu \right) \\ &= \max_{\varphi} \left(-\int \varphi \, d\mu - \int \varphi^* \, d\nu \right). \end{aligned}$$

Here γ ranges over couplings, and the first maximum ranges over proper lower semicontinuous convex functions $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$ satisfying $\varphi(x) + \psi(y) \ge \langle x, y \rangle$ for all x, y. In the second maximum, $\varphi^* \in L^1(\nu)$ is required. The extrema are attained and finite.

Proof. For any admissible pair,

$$\varphi(x) + \psi(y) \ge \langle x, y \rangle. \tag{103}$$

Integrating against any coupling gives

$$\int -\langle x, y \rangle d\gamma \ge -\int \varphi \, d\mu - \int \psi \, d\nu.$$

Taking the infimum over couplings and the supremum over admissible functions proves one inequality. Fenchel's inequality shows that (φ, φ^*) is an admissible pair whenever the required integrability holds.

For the reverse inequality, let $\overline{\gamma}$ be an optimal coupling for the quadratic cost. This is equivalent to minimizing $-\int \langle x, y \rangle d\gamma$, because the integrals of $|x|^2$ and $|y|^2$ depend only on the marginals. Its support is cyclically monotone, so Rockafellar's theorem gives a convex φ with

$$\operatorname{supp}(\overline{\gamma}) \subset \partial\varphi.$$

Consequently,

$$\varphi(x) + \varphi^*(y) = \langle x, y \rangle \quad \overline{\gamma}\text{-almost everywhere.}$$

Both convex functions have affine lower bounds, so their negative parts are integrable under the finite-moment assumptions. The equality above and Cauchy–Schwarz then give integrability of their positive parts as well. Integrating the equality proves duality and attainment. □

Corollary 7.31. *Let μ, ν have finite second moments and let $C(x, y) = |x - y|^2/2$. For a coupling $\overline{\gamma}$, the following are equivalent:*

1. *$\overline{\gamma}$ is optimal;*
2. *$\operatorname{supp}(\overline{\gamma})$ is cyclically monotone;*
3. *there is a proper lower semicontinuous convex function φ such that $\operatorname{supp}(\overline{\gamma}) \subset \partial\varphi$.*

The reverse implication follows by integrating equality in Fenchel's inequality, as in the proof above.

7.8 Brenier's Theorem

Theorem 7.32 (Brenier's Theorem)**.** *Suppose $\int_{\mathbb{R}^n} |x|^2 \, d\mu + \int_{\mathbb{R}^n} |x|^2 \, d\nu < +\infty$, where μ, ν are probability measures and μ is absolutely continuous. Then, there exists a unique optimal plan (coupling) for quadratic cost $\overline{\gamma}$ of μ, ν. Moreover, $\overline{\gamma} = (\mathrm{Id}, T)_{\#}\mu$, i.e., $\overline{\gamma}$ is realized by a transport map unique μ-almost everywhere, and $T = \nabla\varphi$ for some convex function φ.*

Proof. (i) Existence. An optimal plan exists by the Kantorovich existence theorem above.

Suppose $\overline{\gamma}$ is an optimal plan. By Corollary 7.31, we know that $\operatorname{supp}(\overline{\gamma}) \subset \partial\varphi$ for some convex function φ. Also, we have

$$\varphi(x) + \varphi^*(y) = \langle x, y \rangle$$

on $\partial\varphi$ from Kantorovich duality. Thus φ is finite μ-almost everywhere. A convex function is locally Lipschitz, hence differentiable Lebesgue-almost everywhere, in the interior of its domain. The boundary of its full-dimensional convex domain has Lebesgue measure zero.

Since μ is absolutely continuous, φ is therefore differentiable μ-almost everywhere. Let A be the exceptional set. Then, $\mu(A) = 0$ and

$$\operatorname{supp}(\overline{\gamma}) \cap ((\mathbb{R}^n \setminus A) \times \mathbb{R}^n) \subset \{(x, \nabla\varphi(x)) : x \in \mathbb{R}^n\}.$$

Consider any bounded $F \in C(\mathbb{R}^n \times \mathbb{R}^n)$. Then, we have

$$\begin{aligned}\int_{\mathbb{R}^n\times\mathbb{R}^n} F(x,y)\, d\bar{\gamma}(x,y) &= \int_{\mathbb{R}^n\times\mathbb{R}^n} F(x,\nabla\varphi(x))\, d\bar{\gamma}(x,y)\\ &= \int_{\mathbb{R}^n} F(x,\nabla\varphi(x))\, d\mu(x)\\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} F(x,y)\, d((\mathrm{Id},T)_{\#}\mu).\end{aligned}$$

Thus, we can conclude that $\overline{\gamma}$ is realized by optimal mass transport $T = \nabla\varphi$, with convex φ.

(ii) Uniqueness: if γ_1, γ_2 are optimal, then their average is optimal and, by the argument above, is concentrated on the graph of a map. Each γ_i must be concentrated on that same graph. Their first marginals are both μ, so the two plans coincide. The map is unique μ-almost everywhere. □

Corollary 7.33. *Suppose μ, ν are probability measures and μ is absolutely continuous. There exists a map $T_\star\mu = \nu$ of the form $T = \nabla\varphi$ with φ convex. The existence assertion without moment assumptions is the Brenier–McCann extension. For every integrable f,*

$$\int f(y)d\nu(y) = \int f(T(x))d\mu(x).$$

If both measures have densities f_μ, f_ν, the corresponding transport equation is

$$f_\mu(x) = f_\nu(\nabla\varphi(x)) \det \nabla^2\varphi(x) \quad \mu\text{-almost everywhere.}$$

In the smooth case this is the usual change-of-variables formula; in general the Hessian and Jacobian are understood almost everywhere.

7.9 Mass transport proof of the Brunn-Minkowski inequality

Definition 7.34 (Minkowski sum)**.** Suppose K, L are Borel measurable sets in $\mathbb{R}^n$. The *Minkowski sum* of K and L is given as

$$K + L \coloneqq \{x + y : x \in K, y \in L\}.$$

Theorem 7.35 (Brunn-Minkowski inequality)**.** *Suppose K, L are nonempty Borel measurable sets. Then,*

$$|K + L|^{1/n} \geq |K|^{1/n} + |L|^{1/n}, \tag{104}$$

where $|\cdot|$ denotes the volume (Lebesgue measure).

Remark 7.36. *Note that for $\lambda > 0$, $|\lambda K| = \lambda^n |K|$. Thus, we can restate* (104) *as follows: For all $\lambda \in [0,1]$,*

$$|\lambda K + (1-\lambda)L|^{1/n} \geq |\lambda K|^{1/n} + |(1-\lambda)L|^{1/n} = \lambda |K|^{1/n} + (1-\lambda)|L|^{1/n}.$$

That is, $|\cdot|^{1/n}$ is concave with respect to the Minkowski addition.

Remark 7.37. *For $a, b > 0$, $p \in (0,1]$, and $\lambda \in [0,1]$, the weighted arithmetic–geometric mean inequality gives*

$$(\lambda a^p + (1-\lambda)b^p)^{1/p} \geq a^\lambda b^{1-\lambda},$$

(Check for HW). Substituting $p = \frac{1}{n}$, $a = |K|$ and $b = |L|$, we obtain

$$(\lambda |K|^{1/n} + (1-\lambda)|L|^{1/n})^n \geq |K|^\lambda |L|^{1-\lambda}.$$

Under the same assumptions as Theorem 7.35, the Brunn-Minkowski inequality implies

$$|\lambda K + (1-\lambda)L| \geq |K|^\lambda |L|^{1-\lambda} \tag{105}$$

In fact, (105) with all $\lambda \in (0,1)$ implies (104) (Check for HW. Idea: Select right λ and use homogeneity of $|\cdot|$).

Recall the isoperimetric inequality. For the following calculation, take A to be a bounded domain with smooth boundary and positive volume; $|\partial A|$ denotes its surface area. Pick $R > 0$ such that $|A| = |RB_2^n|$. Then, we have

$$|\partial A| \geq |\partial(RB_2^n)|,$$

that is, among all the sets of fixed volume, the Euclidean ball minimizes the perimeter. This is equivalent to saying

$$\frac{|\partial A|}{|A|^{(n-1)/n}} \geq \frac{|\partial(RB_2^n)|}{|RB_2^n|^{(n-1)/n}} = \frac{|\partial B_2^n|}{|B_2^n|^{(n-1)/n}}.$$

Indeed, for all $t > 0$,

$$\frac{|\partial A|}{|A|^{(n-1)/n}} = \frac{t^{n-1}|\partial A|}{t^{n\cdot(n-1)/n}|A|^{(n-1)/n}} = \frac{|\partial(tA)|}{|tA|^{(n-1)/n}}.$$

Also, note that

$$\left|S^{n-1}\right| = n \cdot |B_2^n|,$$

since $|B_2^n| = \int_{B_2^n} dx = \int_0^1 \int_{S^{n-1}} t^{n-1} d\theta dt = |S^{n-1}| \cdot \int_0^1 t^{n-1} dt = \frac{1}{n}|S^{n-1}|$. Therefore, we have

$$\frac{|\partial B_2^n|}{|B_2^n|^{(n-1)/n}} = \frac{n|B_2^n|}{|B_2^n|^{(n-1)/n}} = n|B_2^n|^{1/n},$$

and thus the isoperimetric inequality takes the form

$$\frac{|\partial A|}{|A|^{(n-1)/n}} \geq n\,|B_2^n|^{1/n}. \tag{106}$$

Proof of (106) *via the Brunn-Minkowski inequality.* Note that

$$\begin{aligned}
|\partial A| &= \lim_{\varepsilon \downarrow 0} \frac{|(A+\varepsilon B_2^n) \setminus A|}{\varepsilon} \\
&= \lim_{\varepsilon \downarrow 0} \frac{|A+\varepsilon B_2^n| - |A|}{\varepsilon} && (\partial A \text{ smooth}) \\
&\geq \lim_{\varepsilon \downarrow 0} \frac{\left(|A|^{1/n} + |\varepsilon B_2^n|^{1/n}\right)^n - |A|}{\varepsilon} && (\text{B-M inequality, raised to the power } n) \\
&= n\,|A|^{(n-1)/n}\,|B_2^n|^{1/n}.
\end{aligned}$$

This gives

$$\frac{|\partial A|}{|A|^{(n-1)/n}} \geq n\,|B_2^n|^{1/n} = \frac{|\partial B_2^n|}{|B_2^n|^{(n-1)/n}}.$$

□

For general sets, the same argument directly bounds the outer Minkowski content. The usual finite-perimeter formulation follows by smooth approximation; we keep the smooth notation here for presentation.

Proof of Theorem 7.35 via mass transport. First take bounded sets K, L of positive volume, and let μ, ν be their normalized uniform measures. We write the Jacobian calculation at regular points; the general justification uses the almost-everywhere Jacobian and area formulas for convex gradients. Then

$$\begin{aligned}
d\mu(x) &= \mathbb{1}_K(x)\frac{1}{|K|}dx, \\
d\nu(x) &= \mathbb{1}_L(x)\frac{1}{|L|}dx.
\end{aligned}$$

Consider the Brenier map T such that $T_\star \mu = \nu$. Since $T = \nabla\varphi$ transports μ to ν, the transport equation gives $f_\mu(x) = f_\nu(T(x))\det\nabla^2\varphi(x)$ for almost every $x \in K$. Thus

$$\frac{1}{|K|} = \frac{1}{|L|} \cdot \det(\nabla^2\varphi). \tag{107}$$

Now take $\lambda \in (0,1)$ and let

$$T_\lambda = (1-\lambda)\mathrm{Id} + \lambda T.$$

That is, $x \overset{T_\lambda}{\longmapsto} (1-\lambda)x + \lambda T x$. Then, we have

$$(T_\lambda)_\star \mu = \mu_\lambda,$$

which is concentrated on $(1-\lambda)K + \lambda L$, since $x \in K$ and $T(x) \in L$ for μ-almost every x. Monotonicity of $T = \nabla\varphi$ gives

$$\langle T_\lambda(x) - T_\lambda(y), x - y\rangle \geq (1-\lambda)|x-y|^2,$$

so T_λ is injective and its inverse on its image is Lipschitz. In particular, μ_λ has a density f_λ. The transport equation gives, for almost every $x \in K$,

$$\frac{1}{|K|} = f_\lambda(T_\lambda(x)) \det(\mathrm{Jac}(T_\lambda)(x)).$$

Recall that (104) is equivalent to

$$|(1-\lambda)K + \lambda L| \geq |K|^{1-\lambda} |L|^\lambda . \tag{108}$$

Since $\mu_\lambda((1-\lambda)K + \lambda L) = 1$, it suffices to show, almost everywhere, that

$$f_\lambda(T_\lambda(x)) \leq \frac{1}{|K|^{1-\lambda} |L|^\lambda}. \tag{109}$$

Indeed, if (109) holds, then

$$\int_{(1-\lambda)K+\lambda L} f_\lambda(y)dy \leq |(1-\lambda)K + \lambda L| \operatorname*{ess\,sup}_{x\in K} f_\lambda(T_\lambda(x)) \leq \frac{|(1-\lambda)K + \lambda L|}{|K|^{1-\lambda} |L|^\lambda},$$

while $\int_{(1-\lambda)K+\lambda L} f_\lambda(y)dy = 1$ since f_λ is a density of a probability measure. Thus, (108) and (104) would follow.

Note that (109) is equivalent to

$$\frac{1}{|K| \det(\mathrm{Jac}((1-\lambda)\mathrm{Id} + \lambda T))} \leq \frac{1}{|K|^{1-\lambda} |L|^\lambda},$$

i.e.,

$$\det(\mathrm{Jac}((1-\lambda)\mathrm{Id} + \lambda T)) \geq \left(\frac{|L|}{|K|}\right)^\lambda = \det(\mathrm{Jac}(T))^\lambda. \tag{110}$$

We know that $\mathrm{Jac}(T) = \nabla^2\varphi$, where φ is a convex function. Thus $\mathrm{Jac}(T)$ is a nonnegative definite matrix. Without loss of generality, by choosing the right basis (depending on x), we can assume

$$\mathrm{Jac}(T) = \nabla^2\varphi = \mathrm{diag}(t_1, \ldots, t_n), \qquad t_i \geq 0.$$

Then, (110) reduces to

$$\det(\mathrm{diag}(1-\lambda+\lambda t_1, \cdots, 1-\lambda+\lambda t_n)) \geq \det(\mathrm{diag}(t_1, \cdots, t_n))^\lambda,$$

which is true by the AM-GM inequality ($1-\lambda+\lambda t_i \geq t_i^\lambda$). This proves the result for bounded positive-volume sets. Unbounded sets follow by truncation. If one set has zero volume, nonemptiness implies that the sum contains a translate of the other set, which gives the inequality directly. □

7.10 Log-concave measures

Definition 7.38. A Borel measure μ on $\mathbb{R}^n$ is called *log-concave* if for all Borel measurable sets $K, L \subset \mathbb{R}^n$,

$$\mu(\lambda K + (1-\lambda)L) \geq \mu(K)^\lambda \mu(L)^{1-\lambda},$$

for all $\lambda \in (0,1)$; this is the multiplicative form of the Brunn–Minkowski inequality for μ.

Theorem 7.39 (Borell Theorem)**.** *Suppose $d\mu = f(x)dx$ on $\mathbb{R}^n$, where f is nonnegative, integrable, and log-concave. That is, $\log f$ is concave (with $\log 0 = -\infty$), or equivalently,*

$$f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}$$

for all $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Then, μ is log-concave.

Example 7.40. *The Gaussian measure is log-concave since*

$$\log e^{-|x|^2/2+c} = -\frac{|x|^2}{2} + c$$

is a concave function.

Example 7.41. *Measures with integrable densities of the form $e^{-|x|}$, $e^{-\|x\|}$, or more generally $e^{-(\text{convex function})}$ are log-concave.*

Theorem 7.42 (Prékopa–Leindler inequality)**.** *Let $\lambda \in (0,1)$ and suppose $f, g, h : \mathbb{R}^n \to [0,\infty)$ are measurable, with f, g integrable, and satisfy*

$$h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda},$$

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

$$\int h \geq \left(\int f\right)^\lambda \left(\int g\right)^{1-\lambda}$$

This implies the preceding theorem for Borel sets: if $d\mu = \rho(x)dx$ with ρ log-concave, substitute $f = \mathbb{1}_K \rho$, $g = \mathbb{1}_L \rho$, and $h = \mathbb{1}_{\lambda K + (1-\lambda)L}\rho$.

7.11 Applications of Brenier's theorem in HDP; Caffarelli's theorem

Remark 7.43. *Recall that if μ, ν are probability measures and μ is absolutely continuous, there is a transport $T_\star \mu = \nu$ of the form $T = \nabla\varphi$, where $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex. The pushforward identity is*

$$\int g(y)d\nu(y) = \int g(T(x))d\mu(x).$$

If $d\mu = \rho_\mu(x)dx$ and $d\nu = \rho_\nu(y)dy$, the change-of-variables formula reads

$$\rho_\mu(x) = \rho_\nu(\nabla\varphi(x)) \det \nabla^2\varphi(x) \quad \mu\text{-almost everywhere.}$$

This is the transport equation, or a Monge–Ampère equation. Convexity makes $\nabla^2\varphi$ nonnegative definite.

For presentation, the calculations below are carried out when the densities and transport potential are smooth, the Hessian is positive definite, and the integrations by parts are justified. General versions use smooth approximation and the almost-everywhere transport equation. For the maximum-principle argument, one can use localization or finite differences when a supremum is not attained.

Remark 7.44. *We call a probability measure μ* 1-uniformly log-concave *if*

$$d\mu(x) = e^{-V(x)}dx, \qquad V(x) = \frac{|x|^2}{2} + W(x),$$

where W is convex and the normalizing constant is included in W. In the smooth case this means $\nabla^2 V \geq I_n$.

Gaussian measure conditioned on a convex set K with $\gamma(K) > 0$ is another example:

$$d\mu(x) = \frac{\mathbb{1}_K(x)}{\gamma(K)} d\gamma(x).$$

Here W is an extended-valued convex function,

$$W(x) = \begin{cases} +\infty, & x \notin K, \\ c, & x \in K, \end{cases} \qquad c = \frac{n}{2}\log(2\pi) + \log\gamma(K).$$

Such nonsmooth examples are obtained by approximation.

Theorem 7.45 (Caffarelli's theorem)**.** *Let T be the Brenier map $T_\star\gamma = \mu$, where γ is standard Gaussian and μ is 1-uniformly log-concave. Then T is 1-Lipschitz:*

$$|T(x) - T(y)| \leq |x - y| \qquad \text{for all } x, y \in \mathbb{R}^n.$$

Example 7.46. *(HW) Consider convex bodies $K \subset L$ and their normalized uniform measures $d\mu = |K|^{-1}\mathbb{1}_K dx$, $d\nu = |L|^{-1}\mathbb{1}_L dx$. If $T_\star\mu = \nu$ and T is the Brenier map, T is not necessarily Lipschitz.*[2]

Remark 7.47. *Recall that a 1-Lipschitz transport $T_\star\mu = \nu$ transfers concentration, functional inequalities, and isoperimetric lower bounds from μ to ν.*

[2]**Author query.** Is the intended conclusion that T need not be 1-Lipschitz, or the stronger failure of global Lipschitz continuity? A dilation proves the former; the stronger original wording is retained here pending clarification.

Remark 7.48. *Recall the Gaussian Poincaré inequality: for every sufficiently regular $f : \mathbb{R}^n \to \mathbb{R}$,*

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

Corollary 7.49. *By Remark 7.48 and Caffarelli's theorem, every 1-uniformly log-concave probability measure μ satisfies*

$$\int f^2 d\mu - \left(\int f d\mu\right)^2 \le \int |\nabla f|^2 d\mu$$

for all sufficiently regular $f : \mathbb{R}^n \to \mathbb{R}$.

Theorem 7.50 (A more general form of Caffarelli's theorem)**.** *Let $d\mu = e^{-V(x)}dx$ and $d\nu = e^{-W(x)}dx$ be probability measures, with $V, W \in C^2(\mathbb{R}^n)$ and $\nabla^2 W \ge K I_n$ for some $K > 0$. Let $T = \nabla\varphi$ be the Brenier map from μ to ν. For every unit vector $e \in \mathbb{R}^n$,*

$$\sup_x \varphi_{ee}(x)^2 \le \frac{1}{K} \sup_x V_{ee}(x).$$

Here $f_e = \langle \nabla f, e\rangle$, $f_{ee} = \langle \nabla f_e, e\rangle$, and $f_{ev} = \langle \nabla f_e, v\rangle$ denote directional derivatives. The estimate is understood almost everywhere when the transport potential is not classically smooth.

Example 7.51. *If $\mu = \gamma$, then $V(x) = |x|^2/2 + c_n$, so $V_{ee} = 1$ for every unit vector e. The theorem gives*

$$\sup_x \varphi_{ee}(x)^2 \le \frac{1}{K}.$$

In particular, if $K = 1$, positivity of $\nabla^2\varphi$ implies $0 \le \nabla^2\varphi \le I_n$, so $\nabla\varphi$ is 1-Lipschitz.

Let us sketch the maximum-principle proof in the smooth setting described above. Start with the transport equation

$$e^{-V(x)} = e^{-W(\nabla\varphi(x))} \det \nabla^2\varphi(x).$$

Taking logarithms gives

$$V(x) = W(\nabla\varphi(x)) - \log\det \nabla^2\varphi(x). \tag{111}$$

For a unit vector e, the matrix differentiation formula is

$$\partial_e \log\det \nabla^2\varphi = \operatorname{tr}\big((\nabla^2\varphi)^{-1}\nabla^2\varphi_e\big).$$

Differentiating once more gives

$$\begin{aligned}\partial_{ev} \log\det \nabla^2\varphi = {} & \operatorname{tr}\big((\nabla^2\varphi)^{-1}\nabla^2\varphi_{ev}\big) \\ & - \operatorname{tr}\big((\nabla^2\varphi)^{-1}\nabla^2\varphi_e(\nabla^2\varphi)^{-1}\nabla^2\varphi_v\big).\end{aligned} \tag{112}$$

Combining this with (111),

$$V_{ee} = \langle \nabla^2 W(\nabla\varphi)\nabla^2\varphi e, \nabla^2\varphi e\rangle + \langle \nabla W(\nabla\varphi), \nabla\varphi_{ee}\rangle \\ - \operatorname{tr}\big((\nabla^2\varphi)^{-1}\nabla^2\varphi_{ee}\big) + \operatorname{tr}\Big[\big((\nabla^2\varphi)^{-1}\nabla^2\varphi_e\big)^2\Big].$$

The final term is nonnegative: it equals the trace of the square of the symmetric matrix $(\nabla^2\varphi)^{-1/2}\nabla^2\varphi_e(\nabla^2\varphi)^{-1/2}$.

Suppose φ_{ee} attains its maximum at x_0. Then $\nabla\varphi_{ee}(x_0) = 0$ and $\nabla^2\varphi_{ee}(x_0) \le 0$, so

$$V_{ee}(x_0) \ge \langle \nabla^2 W(\nabla\varphi(x_0))\nabla^2\varphi(x_0)e, \nabla^2\varphi(x_0)e\rangle \ge K|\nabla^2\varphi(x_0)e|^2 \ge K\varphi_{ee}(x_0)^2.$$

Because $\varphi_{ee} \ge 0$, the maximum of φ_{ee} also maximizes its square. Thus

$$\sup_x \varphi_{ee}(x)^2 \le \frac{1}{K}\sup_x V_{ee}(x),$$

as claimed. □

7.12 The Talagrand transport inequality

Theorem 7.52. *Let γ be standard Gaussian measure and μ a probability measure with finite second moment. If $T_\star\gamma = \mu$ is the Brenier map, then*

$$\frac{1}{2}\int |T(x) - x|^2 d\gamma(x) \le \operatorname{Ent}(\mu \parallel \gamma).$$

Here $\operatorname{Ent}(\mu \parallel \gamma) = \int \log(d\mu/d\gamma)d\mu$ *when* $\mu \ll \gamma$*, and is* $+\infty$ *otherwise. The integral on the left is* $W_2(\mu,\gamma)^2/2$*.*

Let us prove this in the smooth setting. If the relative entropy is infinite there is nothing to show. Write

$$g(x) = (2\pi)^{-n/2}e^{-|x|^2/2}, \qquad f(x) = \frac{d\mu}{dx}, \qquad T(x) = \nabla F(x),$$

where F is convex. The transport equation is

$$f(\nabla F(x)) \det \nabla^2 F(x) = g(x).$$

Changing variables $y = \nabla F(x)$ gives

$$\begin{aligned}
\operatorname{Ent}(\mu \parallel \gamma) &= \int f(y)\log\frac{f(y)}{g(y)}dy \\
&= \int f(\nabla F(x))\log\frac{f(\nabla F(x))}{g(\nabla F(x))}\det\nabla^2 F(x)dx \\
&= \int \log\frac{f(T(x))}{g(T(x))}\,d\gamma(x) \\
&= \int \log\frac{g(x)}{g(T(x))\det\nabla^2 F(x)}\,d\gamma(x) \\
&= \int \left(\frac{|T(x)|^2 - |x|^2}{2} - \log\det\nabla^2 F(x)\right) d\gamma(x).
\end{aligned}$$

Comparing this identity with $\frac{1}{2}\int |T(x)-x|^2 d\gamma(x)$, our goal becomes

$$\int \log\det \nabla^2 F \, d\gamma \le \int \big(\langle x, \nabla F\rangle - |x|^2\big) d\gamma.$$

Recall Gaussian integration by parts:

$$\int v \mathrm{L}u \, d\gamma = -\int \langle \nabla v, \nabla u\rangle d\gamma, \qquad \mathrm{L}u = \Delta u - \langle x, \nabla u\rangle.$$

In particular,

$$0 = \int \mathrm{L}F \, d\gamma = \int \big(\Delta F - \langle x, \nabla F\rangle\big) d\gamma, \qquad \int |x|^2 d\gamma = n.$$

It therefore remains to show

$$\int \log\det \nabla^2 F \, d\gamma \le \int (\Delta F - n) d\gamma.$$

This inequality holds pointwise. If $t_1, \ldots, t_n > 0$ are the eigenvalues of $\nabla^2 F$, then

$$\log\det \nabla^2 F = \sum_{i=1}^n \log t_i \le \sum_{i=1}^n (t_i - 1) = \Delta F - n,$$

because $\log t \le t - 1$ for $t > 0$. This proves the theorem. Smooth approximation and lower semicontinuity give the general statement. □

Remark 7.53. *Write now $d\mu = f \, d\gamma$, so f denotes density relative to Gaussian measure. The Gaussian log-Sobolev inequality gives*

$$\mathrm{Ent}(f \, d\gamma \parallel \gamma) \le \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\gamma.$$

The analogous log-Sobolev inequality holds for 1-uniformly log-concave probability measures by Caffarelli's theorem. Combining the Gaussian inequalities yields

$$W_2(\mu, \gamma)^2 \le 2\mathrm{Ent}(\mu \parallel \gamma) \le \int \frac{|\nabla f|^2}{f} d\gamma.$$

References

[1] S. Artstein-Avidan, A. Giannopoulos, V. D. Milman, *Asymptotic Geometric Analysis, part I,* 2015.

[2] N. Alon, B. Klartag, *Optimal compression of approximate inner products and dimension reduction*, Symposium on Foundations of Computer Science (FOCS 2017), 639-650.

[3] Z. D. Bai and Y. Q. Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix,* Ann. Probab. 16 (1988), no. 4, 1729-1741. MR0958213

[4] Z. D. Bai,Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix,* Ann. Probab. 21 (1993), 1275-1294.

[5] J. Bourgain, V. H. Vu and P. M. Wood, *On the singularity probability of discrete random matrices,* J. Funct. Anal. 258 (2010), no. 2, 559-603. MR2557947

[6] N. Cook, *Lower bounds for the smallest singular value of structured random matrices,* Ann. Probab. **46** (2018), no. 6, 3442–3500.

[7] R. Durrett, *Probability: Theory and examples,* 2019.

[8] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. 9 (1988), 543-560.

[9] C. G. Esseen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 5 (1966), 210-216

[10] O. N. Feldheim and S. Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices,* Geom. Funct. Anal. 20 (2010), no. 1, 88-123. MR2647136

[11] Y. Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. Math. 50 (1985), 265–289.

[12] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbb{R}^n$, in Geometric Aspects of Functional Analysis (1986/87), Lecture Notes in Mathematics 1317, Springer, Berlin, 1988, 84–106.

[13] O. Guedon, A. Litvak, K. Tatarko, *Random polytopes obtained by matrices with heavy tailed entries*, Commun. Contemp. Math. **22** (2020), no. 4, 1950027, 28 pp.

[14] R. van Handel, R. Latala, P. Youssef, *The dimension-free structure of nonhomogeneous random matrices*, Invent. Math. **214** (2018), no. 3, 1031–1080.

[15] V. Jain, S. Silwal. *A note on the universality of ESDs of inhomogeneous random matrices*, `arXiv preprint arXiv:2006.05418 (2020).`

[16] J. Kahn, J. Komlos, E. Szemeredi, *On the probability that a random* ± 1 *matrix is singular,* J. Amer. Math. Soc. 8 (1995), 223-240.

[17] R. Kannan, S. Vempala, *Sampling Lattice Points*, Proc. 29th ACM Symposium on the Theory of Computing (STOC '97), El Paso, (1997), Invited for publication in Journal of Comp. and System Sciences.

[18] B. Klartag, *A central limit theorem for convex sets*, Invent. Math., Vol. 168, (2007), 91–131.

[19] B. Klartag, G. V. Livshyts, *The lower bound for Koldobsky's slicing inequality via random rounding*, in Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2017–2019, Volume II, Lecture Notes in Mathematics 2266, Springer, Cham, 2020, 43–63.

[20] B. Klartag, V.D. Milman, *Geometry of log-concave functions and measures*, Geom. Dedicata 112 (2005) 169-182.

[21] J. E. Littlewood, A. C. Offord, *On the number of real roots of a random algebraic equation. III,* Rec. Math. [Mat. Sbornik] N.S. 12 (54), (1943), 277-286

[22] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, Adv. Math. **195** (2005), no. 2, 491–523.

[23] A. E. Litvak and O. Rivasplata, Smallest singular value of sparse random matrices, Studia Math. **212** (2012), no. 3, 195–218.

[24] G. Livshyts, G. Paouris, P. Pivovarov, *Small deviations for operator norms*, work in progress

[25] A. E. Litvak, S. Spektor, *Quantitative version of a Silverstein's result,* GAFA, Lecture Notes in Math., 2116 (2014), 335–340.

[26] G. V. Livshyts, *The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding*, Journal d'Analyse Mathématique 145 (2021), no. 1, 257–306.

[27] G. V. Livshyts, K. Tikhomirov, R. Vershynin, *The smallest singular value of inhomogeneous square random matrices,* Ann. Probab. 49 (2021), no. 3, 1286-1309.

[28] A. Lytova, K. Tikhomirov, *On delocalization of eigenvectors of random non-Hermitian matrices*, Probab. Theory Related Fields **177** (2020), no. 1-2, 465–524.

[29] J. von Neumann, H. H. Goldstine, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. 53 (1947), 1021-1099.

[30] S. Mendelson, G. Paouris, *On the singular values of random matrices,* Journal of the European Mathematics Society, 16, 823-834, 2014.

[31] G. Pólya, G. Szegő, *Aufgaben und Lehrsätze aus der Analysis. Band I: Reihen. Integralrechnung. Funktionentheorie.* Dritte berichtigte Auflage. Die Grundlehren der Mathematischen Wissenschaften, Band 19 Springer-Verlag, Berlin-New York 1964.

[32] P. Raghavan, C. D. Thompson, *Randomized rounding: A technique for provably good algorithms and algorithmic proofs*, Combinatorica 7 (1987), no. 4, 365–374.

[33] E. Rebrova, K. Tikhomirov, *Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries*, Israel J. Math. **227** (2018), no. 2, 507–544.

[34] B. A. Rogozin, *An estimate for the maximum of the convolution of bounded densities*, Teor. Veroyatnost. i Primenen. 32 (1987), no. 1, 53-61, English translation:Theory Probab. Appl. 32 (1987), no. 1, 48-56.

[35] M. Rudelson, *Invertibility of random matrices: norm of the inverse*, Annals of Mathematics 168 (2008), 575-600.

[36] M. Rudelson, R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. 218 (2008), no. 2, 600-633.

[37] M. Rudelson, R. Vershynin, *The least singular value of a random square matrix is* $O(n^{-1/2})$, Comptes Rendus Mathématique 346 (2008), no. 15–16, 893–896.

[38] M. Rudelson, R. Vershynin, *Smallest singular value of a random rectangular matrix,* Communications on Pure and Applied Mathematics 62 (2009), 1707-1739.

[39] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values,* Proceedings of the International Congress of Mathematicians, 2010, pp. 83-120.

[40] M. Rudelson, R. Vershynin, *Small ball probabilities for linear images of high dimensional distributions,* Int. Math. Res. Not. 19 (2015), 9594-9617.

[41] M. Rudelson, R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries,* Duke Math. J. Volume 164, Number 13 (2015), 2507-2538.

[42] G. Schechtman, *A remark concerning the dependence on* ε *in Dvoretzky's theorem*, in Geometric Aspects of Functional Analysis (1987–88), Lecture Notes in Mathematics 1376, Springer, Berlin, 1989, 274–277.

[43] C. Schütt, *Entropy numbers of diagonal operators between symmetric Banach spaces,* J. Approx. Theory 40 (1984), 121–128.

[44] A. Srinivasan, *Approximation Algorithms via Randomized Rounding: a Survey*, Lectures on Approximation and Randomized Algorithms, Series in Advanced Topics in Mathematics, Polish Scientific Publishers PWN, Warsaw, 9-71, (1999).

[45] S. Szarek, *Condition numbers of random matrices, J. Complexity 7 (1991),* 131-149.

[46] K. Tatarko, *An upper bound on the smallest singular value of a square random matrix*, J. Complexity **48** (2018), 119–128.

[47] T. Tao and V. Vu, *On random ±1 matrices: Singularity and Determinant,* Random Structures and Algorithms 28 (2006), no 1, 1-23.

[48] T. Tao, V. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. 20 (2007), 603-628.

[49] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Annals of Mathematics 169 (2009), 595-632.

[50] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values,* Geom. Funct. Anal. 20 (2010), no. 1, 260-297. MR2647142

[51] K. Tikhomirov, *The limit of the smallest singular value of random matrices with i.i.d. entries,* Adv. Math. 284 (2015), 1-20.

[52] K. Tikhomirov, *The smallest singular value of random rectangular matrices with no moment assumptions on entries*, Israel J. Math. 212 (2016), no. 1, 289-314.

[53] K. Tikhomirov, *Invertibility via distance for non-centered random matrices with continuous distributions*, Random Structures & Algorithms 57 (2020), no. 2, 526–562.

[54] K. Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), no. 2, 593–634.

[55] R. Vershynin, *High-dimensional probability: an introduction with applications in data science*, Cambridge University Press, 2018.

[56] R. Vershynin, *Spectral norm of products of random and deterministic matrices,* Probability Theory and Related Fields 150 (2011), 471-509.

[57] M. S. Viazovska, *The sphere packing problem in dimension* 8, Annals of mathematics (2017), 991-1015.

[58] H. Cohn, A. Kumar, S. Miller, D. Radchenko, M. S. Viazovska, *The sphere packing problem in dimension* 24, Annals of mathematics (2017), 185 (3), 1017-1033.

[59] Shiryaev, A. N. *Probability, vol. 1*, Graduate Texts in Mathematics 95 (1989).

[60] S. Jukna, *Extremal Combinatorics: With Applications in Computer Science*, 2nd edition, Texts in Theoretical Computer Science. An EATCS Series, Springer, Berlin, 2011.