Chapter 1

Convexity: the Brunn–Minkowski theory

1.1 Basic facts on convex bodies

We work in the Euclidean space $(\mathbf{R}^n, |\cdot|)$, where $|x| = \left(x_1^2 + \cdots + x_n^2\right)^{1/2}$. We denote by $\langle \cdot\,, \cdot \rangle$ the corresponding inner product. We say that a subset $K \subset \mathbf{R}^n$ is *convex* if for every x, $y \in K$ and $\lambda \in [0,1]$, we have $\lambda x + (1-\lambda) y \in K$. We say that $K \subset \mathbf{R}^n$ is a *convex body* if K is convex, compact, with non-empty interior.

It is convenient to define a distance on the set of convex bodies in $\mathbf{R}^n$. First, given $K \subset \mathbf{R}^n$ and $\varepsilon > 0$, we denote by K_ε the ε-*enlargement* of K, defined as

$$K_\varepsilon = \{x \in \mathbf{R}^n \ : \ \exists y \in K, |x-y| \leqslant \varepsilon\}.$$

In other words, K_ε is the union of closed balls of radius ε with centers in K. The *Hausdorff distance* between two non-empty compact subsets K, $L \subset \mathbf{R}^n$ is then defined as

$$\delta(K, L) = \inf\{\varepsilon > 0 \ : \ K \subset L_\varepsilon \text{ and } L \subset K_\varepsilon\}.$$

We check (check!) that δ is a proper distance on the space of non-empty compact subsets of $\mathbf{R}^n$.

Some basic but important examples of convex bodies in $\mathbf{R}^n$ are

1. The unit Euclidean ball, defined as $B_2^n = \{x \in \mathbf{R}^n \ : \ |x| \leqslant 1\}$.
2. The (hyper)cube $B_\infty^n = [-1,1]^n$.
3. The (hyper)octahedron $B_1^n = \{x \in \mathbf{R}^n \ : \ |x_1| + \cdots + |x_n| \leqslant 1\}$.

These examples are unit balls for the ℓ_p norm on $\mathbf{R}^n$ for $p = 2, \infty, 1$. The ℓ_p norm $\|\cdot\|_p$ is defined for $1 \leqslant p < \infty$ and $x \in \mathbf{R}^n$ by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

and for $p = \infty$ by $\|x\|_\infty = \lim_{p \to \infty} \|x\|_p = \max\{|x_i| \ : \ 1 \leqslant i \leqslant n\}$.

More generally, the following proposition characterizes symmetric convex bodies as the unit balls for some norm.

Proposition 1. *Let $K \subset \mathbf{R}^n$. The following are equivalent*

1. *K is a convex body which is* symmetric *(i.e. satisfies $K = -K$),*
2. *there is a norm on $\mathbf{R}^n$ for which K is the closed unit ball.*

To prove Proposition 1 (check!), we may recover the norm from K by the formula

$$\|x\|_K = \inf\{t > 0 \ : \ \frac{x}{t} \in K\}.$$

A basic geometric fact about convex bodies is given by the Hahn–Banach separation theorem. We give two versions.

Theorem 2. *Let K, L be two convex bodies in $\mathbf{R}^n$ such that $K \cap L = \emptyset$. Then there exist $u \in \mathbf{R}^n$ and $\alpha \in \mathbf{R}$ such that*

$$\max_{x \in K} \langle x, u \rangle < \alpha < \min_{y \in L} \langle y, u \rangle.$$

Here is the geometric meaning of Theorem 2: the hyperplane $H = \{\langle \cdot \, , u \rangle = \alpha\}$ separates K from L, in the sense that each convex body lies in a separate connected component of $\mathbf{R}^n \setminus H$, which is an open half-plane.

Theorem 3. *Let K be a convex body in $\mathbf{R}^n$ and $x \in \partial K$. Then there exists $u \in \mathbf{R}^n$, $u \neq 0$, such that*

$$\max_{y \in K} \langle y, u \rangle = \langle x, u \rangle.$$

The hyperplane $H = \{\langle \cdot \, , u \rangle = \langle x, u \rangle$ is said to be a *support hyperplane* for K at the boundary point x. One can give a geometric proof of Theorem 2 (check!) as follows: choose a couple of points $(x, y) \in K \times L$ which minimizes $|x - y|$, and take as a separating hyperplane the set of points equidistant from x and y. We can then obtain Theorem 3 as a corollary by separating K from $\{x_k\}$, where (x_k) is a sequence in $\mathbf{R}^n \setminus K$ converging to x (check!).

1.2 The Brunn–Minkowski inequality

Given sets K, L in $\mathbf{R}^n$ and a nonzero real number λ, we may define

$$\lambda K = \{\lambda x \ : \ x \in K\},$$

$$K + L = \{x + y \ : \ x \in K, y \in L\},$$

which we call the Minkowski sum of K and L. We denote by $\mathrm{vol}(\,\cdot\,)$ the Lebesgue measure, or volume, defined on Borel subsets of $\mathbf{R}^n$. We may write vol_n instead of vol if we want to precise the dimension. The volume is n-homogeneous, i.e. satisfies $\mathrm{vol}(\lambda A) = |\lambda|^n \mathrm{vol}(A)$, for $\lambda \in \mathbf{R}$. The behaviour of the volume with respect to Minkowski addition is governed by the Brunn–Minkowski inequality.

Theorem 4 (Brunn–Minkowski inequality). *Let K, L be compact subsets of $\mathbf{R}^n$, and $\lambda \in (0,1)$. Then*

$$\mathrm{vol}(\lambda K + (1-\lambda)L) \geqslant \mathrm{vol}(K)^{\lambda}\,\mathrm{vol}(L)^{1-\lambda}. \tag{1.1}$$

In other words, the function $\log \mathrm{vol}$ is concave with respect to Minkowski addition. Before proving the Brunn–Minkowski inequality, we point that there is an equivalent form: for every nonempty compact sets A, B in $\mathbf{R}^n$, we have

$$\mathrm{vol}(A+B)^{1/n} \geqslant \mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}. \tag{1.2}$$

We check the equivalence between (1.1) and (1.2) by taking advantage of the homogeneity of the volume. To show (1.2) from (1.1), consider the numbers $a = \mathrm{vol}(A)^{1/n}$ and $b = \mathrm{vol}(B)^{1/n}$. The case when $a = 0$ (and, similarly, $b = 0$) is easy: it suffices to notice that $A + B$ contains a translate of A (check!). If $ab > 0$, we may write

$$A + B = (a+b)\left[\frac{a}{a+b}\frac{A}{a} + \frac{b}{a+b}\frac{B}{b}\right],$$

and conclude from (1.1) that $\mathrm{vol}(A+B) \geqslant (a+b)^n$, as needed. For the converse implication, we write

$$\begin{aligned}
\mathrm{vol}(\lambda K + (1-\lambda)L)^{1/n} &\geqslant \mathrm{vol}(\lambda K)^{1/n} + \mathrm{vol}((1-\lambda)L)^{1/n} \\
&= \lambda\,\mathrm{vol}(K)^{1/n} + (1-\lambda)\,\mathrm{vol}(L)^{1/n} \\
&\geqslant \left[\mathrm{vol}(K)^{1/n}\right]^{\lambda}\left[\mathrm{vol}(L)^{1/n}\right]^{1-\lambda},
\end{aligned}$$

where the last step is the arithmetic mean-geometric mean (AM-GM) inequality (check!).

We present the proof of a functional version of the Brunn–Minkowski inequality.

Theorem 5 (Prékopa–Leindler inequality)**.** *Let $\lambda \in (0,1)$. Assume that f, g, $h : \mathbf{R}^n \to [0,\infty]$ are measurable functions such that,*

$$\text{for every } x, y \in \mathbf{R}^n, \quad h(\lambda x + (1-\lambda)y) \geqslant f(x)^\lambda g(y)^{1-\lambda}. \tag{1.3}$$

Then,

$$\int_{\mathbf{R}^n} h \geqslant \left(\int_{\mathbf{R}^n} f \right)^\lambda \left(\int_{\mathbf{R}^n} g \right)^{1-\lambda}.$$

Before proving Theorem 5, we notice that it immediately implies Theorem 4 by choosing $f = \mathbf{1}_K$, $g = \mathbf{1}_L$ and $h = \mathbf{1}_{\lambda K + (1-\lambda)L}$.

Proof of Theorem 5. The proof is by induction on the dimension n. We first consider the base case $n = 1$. By monotone convergence, we may reduce to the case where f, g are bounded, and by homogeneity to the case when $\|f\|_\infty = \|g\|_\infty = 1$ (check!). We also use the following formula which relates integrals with measures of level sets (check!): whenever $\phi : X \to \mathbf{R}^n$ is a measurable function defined on a measure space (X, μ), then

$$\int_X \phi = \int_0^\infty \mu(\{\phi \geqslant t\}) \, \mathrm{d}t. \tag{1.4}$$

Another information we need is that the Brunn–Minkowski inequality holds in dimension 1: for nonempty measurable sets A, B in $\mathbf{R}$ such that $A + B$ is measurable, we have $\operatorname{vol}(A+B) \geqslant \operatorname{vol}(A) + \operatorname{vol}(B)$. To prove this, reduce to the case when $\sup A < +\infty$ and $\inf B > -\infty$, and show that $A + B$ contains disjoint translates of A and B (check!).

The proof goes as follows: for $0 \leqslant a < 1$, we have

$$\{h \geqslant a\} \supset \lambda\{f \geqslant a\} + (1-\lambda)\{g \geqslant a\},$$

which by the one-dimensional Brunn–Minkowski implies

$$\operatorname{vol}(\{h \geqslant a\}) \geqslant \lambda \operatorname{vol}(\{f \geqslant a\}) + (1-\lambda) \operatorname{vol}(\{g \geqslant a\}).$$

We then integrate this inequality when a ranges over $[0,1)$, and use (1.4) 3 times to obtain

$$\begin{aligned} \int_{\mathbf{R}} h &\geqslant \lambda \int_{\mathbf{R}} f + (1-\lambda) \int_{\mathbf{R}} g \\ &\geqslant \left(\int_{\mathbf{R}^n} f \right)^\lambda \left(\int_{\mathbf{R}^n} g \right)^{1-\lambda} \end{aligned}$$

by the AM-GM inequality.

We now explain the induction step, assuming the result in dimension n. We decompose $\mathbf{R}^{n+1}$ as $\mathbf{R}^n \times \mathbf{R}$. Let $f, g, h : \mathbf{R}^{n+1} \to \mathbf{R}$ satisfying (1.3). For $y \in \mathbf{R}$, we define 3 functions on $\mathbf{R}^n$ by the formulas $f_y(t) = f(t,y)$, $g_y(t) = g(t,y)$, $h_y(t) = h(t,y)$. Whenever real numbers y, y_1, y_2 are such that $y = \lambda y_1 + (1-\lambda)y_2$, we have $h_y(\lambda s_1 + (1-\lambda)s_2) \geqslant f_{y_1}(s_1)^\lambda g_{y_2}(s_2)^{1-\lambda}$ for $s_1, s_2 \in \mathbf{R}^n$. In other words, the functions f_{y_1}, g_{y_2}, h_y satisfy the hypothesis (1.3). By the induction step, it follows that

$$\int_{\mathbf{R}^n} h_y \geqslant \left(\int_{\mathbf{R}^n} f_{y_1}\right)^\lambda \left(\int_{\mathbf{R}^n} g_{y_2}\right)^{1-\lambda}.$$

If we define functions F, G, H on $\mathbf{R}$ by $F(y) = \int_{\mathbf{R}^n} f_y$, $G(y) = \int_{\mathbf{R}^n} g_y$ and $H(y) = \int_{\mathbf{R}^n} h_y$, this means that the functions F, G, H also satisfy (1.3). By using the case $n = 1$, and Fubini theorem, it follows that

$$\int_{\mathbf{R}^{n+1}} h \geqslant \left(\int_{\mathbf{R}^{n+1}} f\right)^\lambda \left(\int_{\mathbf{R}^{n+1}} g\right)^{1-\lambda}. \qquad \square$$

A remarkable corollary of the Brunn–Minkowski theorem is the *isoperimetric inequality.* One may define the surface area of a subset $K \subset \mathbf{R}^n$ by

$$a(K) = \limsup_{\varepsilon \to 0} \frac{\mathrm{vol}(K_\varepsilon) - \mathrm{vol}(K)}{\varepsilon}. \tag{1.5}$$

This is a simple way to define the $(n-1)$-dimensional measure of ∂K.

Theorem 6 (Isoperimetric inequality)**.** *Let $K \subset \mathbf{R}^n$ be a compact set with $\mathrm{vol}(K) > 0$, and B a Euclidean ball with radius chosen so that $\mathrm{vol}(K) = \mathrm{vol}(B)$. Then, for every $\varepsilon > 0$, we have $\mathrm{vol}(K_\varepsilon) \geqslant \mathrm{vol}(B_\varepsilon)$, and therefore $a(K) \geqslant a(B)$.*

Proof. We may take $B = rB_2^n$, for $r = \left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(B_2^n)}\right)^{1/n}$. We have then $B_\varepsilon = (r+\varepsilon)B_2^n$. Note that $K_\varepsilon = K + \varepsilon B_2^n$. By (1.2), we have

$$\begin{aligned} \mathrm{vol}(K_\varepsilon)^{1/n} &\geqslant \mathrm{vol}(K)^{1/n} + \mathrm{vol}(\varepsilon B_2^n)^{1/n} \\ &= (r+\varepsilon)\,\mathrm{vol}(B_2^n)^{1/n} \\ &= \mathrm{vol}(B_\varepsilon) \end{aligned}$$

as needed. $\square$

Theorem 6 can be rephrased as follows: at fixed volume, Euclidean balls minimize the surface area.

1.3 The Blaschke–Santalò inequality

We introduce now polarity. The *polar* of a set $K \subset \mathbf{R}^n$ is defined as

$$K^\circ = \{x \in \mathbf{R}^n \ : \ \forall y \in K, \langle x, y \rangle \leqslant 1\}.$$

We emphasize that polarity depends on the choice of a inner product. Polarity at the level of unit balls corresponds to duality for normed spaces. Indeed, given a norm $\|\cdot\|$ on $\mathbf{R}^n$, we may (using the standard inner product of $\mathbf{R}^n$) identify the normed space dual to $(\mathbf{R}^n, \|\cdot\|)$ with $(\mathbf{R}^n, \|\cdot\|_*)$. If K is the unit ball for $\|\cdot\|$, then (check!) K° is the unit ball for $\|\cdot\|_*$.

We list basic properties of polarity (check!)

- If K is a symmetric convex body, then $(K^\circ)^\circ = K$, a statement known as the bipolar theorem.
- $(B_1^n)^\circ = B_\infty^n$, $(B_2^n)^\circ = B_2^n$ and $(B_\infty^n)^\circ = B_1^n$.
- If $K \subset L$, then $K^\circ \supset L^\circ$.
- Whenever $T \in \mathsf{GL}_n(\mathbf{R})$ is an invertible linear map, then $T(K)^\circ = (T^*)^{-1}(K^\circ)$, where T^* is the transpose (or adjoint) of T. In particular, $(\alpha K)^\circ = \alpha^{-1} K^\circ$ whenever $\alpha \in \mathbf{R}^*$.

A consequence of the last property is that, for K a convex body and $T \in \mathsf{GL}_n(\mathbf{R})$,

$$\operatorname{vol}(TK)\operatorname{vol}((TK)^\circ) = \operatorname{vol}(K)\operatorname{vol}(K^\circ).$$

In other words, the quantity $\operatorname{vol}(K)\operatorname{vol}(K^\circ)$, sometimes called the volume product of K, is invariant under the action of the linear group. The Blaschke–Santalò inequality shows that, among symmetric convex bodies, this quantity is maximal for the Euclidean ball.

Theorem 7 (Blaschke–Santalò inequality). *If $K \subset \mathbf{R}^n$ is a symmetric convex body, then*

$$\operatorname{vol}(K)\operatorname{vol}(K^\circ) \leqslant \operatorname{vol}(B_2^n)^2.$$

We will present a proof of the Blaschke–Santalò by symmetrization: we explicit a geometric process which bring any symmetric body "closer" to the Euclidean ball, while increasing the volume product.

Given a convex body $K \subset \mathbf{R}^n$ and a direction $u \in S^{n-1}$ (the unit sphere), we define the Steiner symmetrization of K in the direction u, denoted $S_u K$, as follows. For every $x \in u^\perp$, we define

$$S_u K \cap (x + \mathbf{R}u) = \begin{cases} \emptyset & \text{if } K \cap (x+\mathbf{R}u) = \emptyset \\ [x - \frac{\alpha}{2}u, x + \frac{\alpha}{2}u] & \text{otherwise, where } \alpha = \operatorname{vol}_1(K \cap (x + \mathbf{R}u)). \end{cases}$$

The geometric meaning is the following: we write K as a union of segments parallel to u, and translate each of these segments along u such that each midpoint belongs to the hyperplane $u^\perp$. One may check (check!) the formula

$$S_u K = \left\{ x + \frac{s-t}{2} u \ : \ x \in u^\perp, s,t \in \mathbf{R} \text{ are such that } x + su \in K \text{ and } x + tu \in K \right\}.$$

Some properties of the Steiner symmetrization are

- It preserves volume: $\mathrm{vol}(S_u K) = \mathrm{vol}(K)$, as a consequence of Fubini theorem (check!).
- It is increasing: $K \subset L$ implies $S_u K \subset S_u L$.
- It preserves convexity: if K is a convex body, then $S_u K$ is a convex body, as a consequence of the 1-dimensional Brunn–Minkowski inequality (check!).

In order to prove Blaschke–Santalò inequality using Steiner symmetrizations, we are going to need more sophisticated properties.

Proposition 8. *If $K \subset \mathbf{R}^n$ is a symmetric convex body, then for every $u \in S^{n-1}$,*

$$\mathrm{vol}(K^\circ) \leqslant \mathrm{vol}((S_u K)^\circ))$$

and therefore $\mathrm{vol}(K)\,\mathrm{vol}(K^\circ) \leqslant \mathrm{vol}(S_u K)\,\mathrm{vol}((S_u K)^\circ))$.

Proposition 9. *Let $K \subset \mathbf{R}^n$ be a symmetric convex body, and denote by $\mathcal{A}$ the set of convex bodies obtained by applying to K finitely many Steiner symmetrizations, in any directions. Then there is a sequence (K_k) in $\mathcal{A}$ which converges, in Hausdorff distance, towards rB_2^n, where $r = \left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(B_2^n)}\right)^{1/n}$.*

In order to derive Theorem 7 for Propositions 8 and 9, it suffices to check that the function $L \mapsto \mathrm{vol}(L^\circ)$ (defined on the set of symmetric convex bodies) is continuous for the Hausdorff distance (check!).

Proof of Proposition 8. Without loss of generality (check!), we may assume that $u = (0, \dots, 0, 1)$. We identify $\mathbf{R}^n$ with $\mathbf{R}^{n-1} \times \mathbf{R}$. We have

$$S_u K = \left\{ \left(x, \frac{s-t}{2}\right) \ : \ (x,s) \in K, (x,t) \in K \right\},$$

$$(S_u K)^\circ = \left\{ (y,r) \ : \ \langle x, y \rangle + \frac{r(s-t)}{2} \leqslant 1 \ \ \forall (x,s), (x,t) \in K \right\}.$$

We use the following notation: given $A \subset \mathbf{R}^n$ and $r \in \mathbf{R}$, we set $A[r] = \{x \in \mathbf{R}^{n-1} \ : (x,r) \in A\}$. We claim that

$$\frac{1}{2}\left(K^\circ[r] + K^\circ[-r]\right) \subset (S_u K)^\circ[r]. \tag{1.6}$$

The left hand-side of (1.6) is equal to

$$\left\{ \frac{y+z}{2} \ : \ \langle y, x \rangle + rs \leqslant 1 \text{ and } \langle z, w \rangle - rt \leqslant 1 \quad \text{whenever } (x,s) \in K, (w,t) \in K \right\},$$

which is a subset of (we have a larger set since we ask for fewer constraints by requiring $w = x$)

$$\left\{ \frac{y+z}{2} \ : \ \langle y, x\rangle + rs \leqslant 1 \text{ and } \langle z, x\rangle - rt \leqslant 1 \quad \text{whenever } (x,s) \in K, (x,t) \in K \right\},$$

and further a subset of (requiring the sum of two inequality is true is less demanding than requiring each inequality)

$$\left\{ v \ : \ \langle v, x\rangle + \frac{(s-t)r}{2} \leqslant 1 \quad \text{whenever } (x,s) \in K, (x,t) \in K \right\},$$

which is the right hand-side of (1.6).

Since K a symmetric convex body, we have $K^\circ[r] = -K^\circ[-r]$. In particular, this implies that $\mathrm{vol}(K^\circ[r]) = \mathrm{vol}(K^\circ[-r])$. By the Brunn–Minkowski inequality, we have therefore $\mathrm{vol}((S_u K)^\circ[r]) \geqslant \mathrm{vol}(K^\circ[r])$. Since this holds for every $r \in \mathbf{R}$, we obtain the inequality $\mathrm{vol}((S_u K)^\circ) \geqslant \mathrm{vol}(K^\circ)$ using the Fubini theorem. □

The proof of Proposition 9 uses a compactness argument on the set of convex bodies, which is most easily discussed in terms of *support functions*. The support function of a convex body K is the function $h_K : \mathbf{R}^n \to \mathbf{R}$ defined as

$$h_k(u) = \max_{x \in K} \langle x, u\rangle.$$

If K is a symmetric convex body, then h_K coincides with $\|\cdot\|_{K^\circ}$, the norm for which K° is the unit ball. Some properties of the support function are (for convex bodies K, L),

- $K \subset L$ if and only if $h_K \leqslant h_L$ (check!),
- we have the identity (check!)

$$\delta(K, L) = \sup_{u \in S^{n-1}} |h_K(u) - h_L(u)|, \tag{1.7}$$

- we can also recover K from h_K by the formula (check!)

$$K = \bigcap_{u \in S^{n-1}} \{\langle \cdot\, , u\rangle \leqslant h_K(u)\}. \tag{1.8}$$

Theorem 10 (Blaschke selection theorem)**.** *Let (K_k) be a sequence of convex bodies satisfying $rB_2^n \subset K_k \subset RB_2^n$ for some r, R. Then there exists a subsequence of (K_k) which converges in Hausdorff distance to a convex body K.*

Proof. Consider the family of functions h_{K_k}, seen as a subset of the Banach space $C(S^{n-1})$ of continuous functions on the sphere, equipped with the sup norm. For every k, the function h_{K_k} is R-Lipschitz (check!). By Ascoli's theorem, it follows that some subsequence converges uniformly on S^{n-1} to a function $h \in C(S^{n-1})$. The last step is to show that we can find a convex body K such that $h = h_K$ (check!using formula (1.8)). By (1.7), uniform convergence of the support functions towards h_K is equivalent to convergence towards K in Hausdorff distance. □

Proof of Proposition 9. We denote by $\overline{\mathcal{A}}$ the closure of $\mathcal{A}$ (inside the space of all convex bodies) with respect to Hausdorff distance. Using Blaschke selection theorem, we check that the continuous function

$$L \mapsto \mathrm{vol}(L \cap rB_2^n)$$

achieves its maximum on $\mathcal{A}$, say at L_0. Assume now that $\mathrm{vol}(L_0 \cap rB_2^n) < \mathrm{vol}(rB_2^n)$. Then there exist $x \in rB_2^n \setminus L_0$ and $y \in L_0 \setminus rB_2^n$. Define now $u = \frac{x-y}{|x-y|} \in S^{n-1}$. We check that (check! – consider the line going through x and y)

$$S_u(L_0 \cap rB_2^n) \subsetneq S_u(L_0) \cap rB_2^n,$$

and therefore $\mathrm{vol}(S_u(L_0 \cap rB_2^n)) < \mathrm{vol}(S_u(L_0))$, contradicting the maximality of L_0 (check! – use the fact that volume for convex bodies is continuous with respect to the Hausdorff distance). It follows that $\mathrm{vol}(L_0 \cap rB_2^n) = \mathrm{vol}(rB_2^n)$, and therefore $L_0 = rB_2^n$. □

Finally we mention the following conjecture

Conjecture 11 (Mahler)**.** *If $K \subset \mathbf{R}^n$ is a symmetric convex body, then*

$$\mathrm{vol}(K)\,\mathrm{vol}(K^\circ) \geqslant \mathrm{vol}(B_1^n)\,\mathrm{vol}(B_\infty^n).$$

Mahler's conjecture has been proved only in dimensions 2 and 3.

Chapter 2

The Banach–Mazur compactum

2.1 Banach–Mazur distance, ellipsoids

In this chapter we study the set of normed spaces of dimension n. Any such space is isometric to $(\mathbf{R}^n, \| \cdot \|)$ for some norm. The choice of norm is not unique: for any $T \in \mathsf{GL}_n(\mathbf{R})$, the normed spaces $X_1 = (\mathbf{R}^n, \| \cdot \|)$ and $X_2 = (\mathbf{R}^n, \|T(\cdot)\|)$ are isometric. If K is the unit ball for X_1, then $T^{-1}(K)$ is the unit ball for X_2. Studying n-dimensional normed spaces up to isometry is equivalent to studying symmetric convex bodies in $\mathbf{R}^n$ up to the action of $\mathsf{GL}_n(\mathbf{R})$.

If X and Y are n-dimensional normed space, define their *Banach–Mazur distance* as

$$d_{BM}(X,Y) = \inf \left\{ \|T : X \to Y\| \cdot \|T^{-1} : Y \to X\| \ : \ T : X \to Y \text{ linear bijection} \right\}.$$

Here $\|T : X \to T\|$ is the *operator norm* of T, i.e. $\sup\{\|Tx\|_Y \ : \ \|x\|_X \leqslant 1\}$. At the level of unit balls (denoted B_X and B_Y, the quantity $\|T : X \to Y\|$ is the smallest $\lambda \geqslant 0$ such that $T(B_X) \subset \lambda B_Y$.

We define similarly the Banach–Mazur distance between two symmetric convex bodies K, $L \subset \mathbf{R}^n$

$$d_{BM}(K,L) = \inf \left\{ \frac{b}{a} \ : \ aK \subset T(L) \subset bK \text{ for } a,b > 0 \text{ and } T \in \mathsf{GL}_n(\mathbf{R}) \right\}.$$

Here are some basic properties of d_{BM}. We note that $\log d_{BM}$ satisfies the axioms of a distance.

- symmetry: we have $d_{BM}(K,L) = d_{BM}(L,K)$ because $aK \subset T(L) \subset bK$ is equivalent to $b^{-1}L \subset T^{-1}(K) \subset a^{-1}L$.
- invariance under polarity: we have $d_{BM}(K,L) = d_{BM}(K^\circ, L^\circ)$ because $aK \subset T(L) \subset bK$ is equivalent to $b^{-1}K^\circ \subset (T^*)^{-1}(L^\circ) \subset a^{-1}K^\circ$.

- triangular inequality: we have $d(K,M) \leqslant d(K,L)d(L,M)$
- $d(K,L) = 1$ is equivalent to the fact that there is $T \in \mathsf{GL}_n(\mathbf{R})$ such that $T(K) = L$ (check! using compactness).

We denote by BM_n the set of symmetric convex bodies in $\mathbf{R}^n$, up to the equivalence relation

$$K \sim L \iff \exists T \in \mathsf{GL}_n(\mathbf{R}) \ : \ L = T(K).$$

The space $(BM_n, \log d_{BM})$ is a metric space. As we will see later, it is compact and often called *the Banach–Mazur compactum.*

An *ellipsoid* $\mathcal{E} \subset \mathbf{R}^n$ is a convex body of the form $\mathbf{E} = T(B_2^n)$ for $T \in \mathsf{GL}_n(\mathbf{R})$. We first give a characterization of ellipsoids. We denote by M_n^+ (resp. M_n^{++} the cone of nonnegative (resp. positive) $n \times n$ symmetric matrices.

Proposition 12. *For $\mathcal{E} \subset \mathbf{R}^n$, the following are equivalent*

1. *$\mathcal{E}$ is an ellipsoid,*
2. *there is a $A \in \mathsf{M}_n^{++}$ such that $\mathcal{E} = A(B_2^n)$,*
3. *there is an orthonormal basis (f_i) of $\mathbf{R}^n$, and positive numbers (α_i), such that*

$$\mathcal{E} = \left\{ x \in \mathbf{R}^n \ : \ \sum_{i=1}^n \alpha_i^{-2} \langle x, f_i \rangle^2 \leqslant 1 \right\}.$$

4. *There is a inner product on $\mathbf{R}^n$ such that $\mathcal{E}$ is the unit ball for the associated norm.*

Proof. The equivalence between 1. and 2. follows from the polar decomposition: any $T \in \mathsf{GL}_n(\mathbf{R})$ can be written as $T = AO$ for $O \in \mathsf{O}(n)$ and $A \in \mathsf{M}_n^{++}$. We then have $T(B_2^n) = A(O(B_2^n)) = A(B_2^n)$.

To show that 2. implies 3., use the spectral theorem to diagonalize A in an orthonormal basis (f_i), i.e. $Af_i = \alpha_i f_i$ for $\alpha_i > 0$. For $x \in \mathbf{R}^n$, we have $A(x) = \sum_i \alpha_i \langle x, f_i \rangle f_i$ and $A^{-1}(x) = \alpha_i^{-1} \langle x, f_i \rangle f_i$. It follows that

$$x \in \mathcal{E} \iff A^{-1}(x) \in B_2^n \iff \sum_i \alpha_i^{-2} \langle x, f_i \rangle^2 \leqslant 1.$$

To get 4. from 3., consider the inner product

$$Q(x,y) = \sum_{i=1}^n \alpha_i^{-2} \langle x, f_i \rangle \langle y, f_i \rangle.$$

To get 2. from 4., use the fact that any inner product Q can be written as $Q(x,y) = \langle x, Ax \rangle$ for a positive matrix A. It follows that

$$Q(x,x) \leqslant 1 \iff \langle x, Ax \rangle \leqslant 1 \iff |A^{1/2}x| \leqslant 1 \iff x \in A^{-1/2}(B_2^n). \qquad \square$$

2.2 John's theorem

John's theorem allows to estimate the Banach–Mazur distance between B_2^n and an arbitrary convex body.

We use the following notation: given x, $y \in \mathbf{R}^n$, we denote by $|x\rangle\langle y|$ the linear map (of rank 1) given by $z \mapsto \langle y, z\rangle x$ (a linear map from $\mathbf{R}^n$ to $\mathbf{R}^n$). In terms of matrices, this correspond to the matrix $(x_i y_j)_{1\leqslant i,j\leqslant n}$.

Theorem 13 (John's theorem)**.** *Let $K \subset \mathbf{R}^n$ be a symmetric convex body. Then there is a unique ellipsoid of maximal volume inside K, denoted $\mathcal{E}_J(K)$ and called the* John ellipsoid *of K. Moreover, we have the equivalence*

$$\mathcal{E}_J(K) = B_2^n \iff B_2^n \subset K \text{ and } \frac{\mathrm{Id}}{n} \in \operatorname{conv}\{|x\rangle\langle x| \ : \ x \in \partial K \cap S^{n-1}\}.$$

Intuitively, if $B_2^n \subset K$ but without enough "contact point", then there is a way to construct another ellipsoid inside K with a larger volume. When $\mathcal{E}_J(K) = B_2^n$, we say that K is in the *John position.* For every symmetric convex body $K \subset \mathbf{R}^n$, there is $T \in \mathsf{GL}_n(\mathbf{R})$ such that $T(K)$ is in the John position.

We first look at two examples. Note that the inclusions $B_1^n \subset B_2^n \subset B_\infty^n$ and $\frac{1}{\sqrt{n}}B_\infty^n \subset B_2^n \subset \sqrt{n}B_1^n$ are sharp.

1. The John ellipsoid of B_∞^n is B_2^n. This is because we have $\frac{\mathrm{Id}}{x} = \sum_{i=1}^n |e_i\rangle\langle e_i|$, where (e_i) is the canonical basis.

2. The John ellipsoid of B_1^n is $\frac{1}{\sqrt{n}}B_2^n$, or equivalently the John ellipsoid of $\sqrt{n}B_1^n$ is B_2^n. What is the set $\sqrt{n}B_1^n \cap S^{n-1}$? This contains elements x such that $\sum x_i^2 = 1$ and $\sum |x_i| = \sqrt{n}$. Using the equality case in the Cauchy–Schwarz inequality $\sum |x_i| \leqslant \sqrt{n}\sum x_i^2$, we check that $\sqrt{n}B_1^n \cap S^{-1} = \{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$. If x is uniformly distributed on this set, we have (check!)
$$\mathbf{E}\, |x\rangle\langle x| = \frac{\mathrm{Id}}{n}.$$

Proof of John's theorem. We first show existence. We note that if $\mathcal{E} = TB_2^n$, then $\operatorname{vol}(\mathcal{E}) = |\det(T)| \operatorname{vol}(B_2^n)$. The set
$$\{T \in \mathsf{M}_n(\mathbf{R}) \ : \ T(B_2^n) \subset K\}$$
is compact (check!) and therefore the continuous function $|\det(\,\cdot\,)|$ achieves its maximum.

For the uniqueness, we use the following lemma.

Lemma 14. *The function* $\log\det$ *is strictly concave on* M_n^{++}.

Proof. For T_1, $T_2 \in \mathsf{M}_n^{++}$, we have

$$\det\left(\frac{T_1 + T_2}{2}\right) = \det(T_1)\det\left(\frac{\mathrm{Id} + T_1^{-1/2}T_2T_1^{-1/2}}{2}\right).$$

If we denote $A = T_1^{-1/2} T_2 T_1^{-1/2} \in \mathsf{M}_n^{++}$, then $\det(A) = \det(T_2)/\det(T_1)$. Let (λ_i) be the eigenvalues of A. By the concavity of log, we have

$$\log\det\left(\frac{\mathrm{Id}+A}{2}\right) = \log\prod_{i=1}^n \left(\frac{1+\lambda_i}{2}\right) = \sum_{i=1}^n \log\left(\frac{1+\lambda_i}{2}\right) \geqslant \frac{1}{2}\sum_{i=1}^n \log\lambda_i = \frac{1}{2}\log\frac{\det(T_2)}{\det(T_1)}.$$

It follows that

$$\det\left(\frac{T_1+T_2}{2}\right) = \det(T_1)\det\left(\frac{\mathrm{Id}+A}{2}\right) \geqslant \sqrt{\det(T_1)\det(T_2)}.$$

Moreover, since log is strictly concave, there is equality if and onlf if $\lambda_i = 1$ for every i, i.e. $T_1 = T_2$. □

We now prove uniqueness in John's theorem. Suppose that $\mathcal{E}_1$, $\mathcal{E}_2$ are two ellipsoids inside K with $\mathrm{vol}(\mathcal{E}_1) = \mathrm{vol}(\mathcal{E}_2)$. We may write $\mathcal{E}_1 = T_1(B_2^n)$ and $\mathcal{E}_2 = T_2(B_2^n)$ for T_1, $T_2 \in \mathsf{M}_n^{++}$. Necessarily $\det(T_1) = \det(T_2)$. Consider the ellipsoid $\mathcal{E} = \left(\frac{T_1+T_2}{2}\right)(B_2^n)$. By the previous lemma, it satisfies $\mathrm{vol}(\mathcal{E}) > \mathrm{vol}(\mathcal{E}_1)$ while $\mathcal{E} \subset \frac{\mathcal{E}_1+\mathcal{E}_2}{2} \subset K$.

We now prove the characterization. First assume that $B_2^n \subset K$, and that there exist contact points (x_i) in $\partial K \cap S^{n-1}$ and a convex combination (λ_i) such that $\sum \lambda_i |x_i\rangle\langle x_i| = \mathrm{Id}/n$. It follows that for every y, z in $\mathbf{R}^n$,

$$\frac{\langle y, z\rangle}{n} = \sum \lambda_i \langle y, x_i\rangle\langle x_i, z\rangle. \tag{2.1}$$

Consider an ellipsoid $\mathcal{E} \subset K$, of the form

$$\mathcal{E} = \{x \in \mathbf{R}^n \ : \ \sum \alpha_j^{-2}\langle x, f_j\rangle^2 \leqslant 1\}$$

for an orthonormal basis (f_j). It follows (check!) that

$$\mathcal{E}^\circ = \{x \in \mathbf{R}^n \ : \ \sum \alpha_j^{2}\langle x, f_j\rangle^2 \leqslant 1\}.$$

For every i, since $x_i \in \partial K \cap S^{n-1}$, it must be (check!) that $K \subset \{y \ : \ \langle y, x_i\rangle \leqslant 1\}$, so that $x_i \in K^\circ \subset \mathcal{E}^\circ$ and therefore $\sum_j \alpha_j^2 \langle x_i, f_j\rangle^2 \leqslant 1$. Taking convex combinations gives

$$\sum_i \lambda_i \sum_j \alpha_j^2 \langle x_i, f_j\rangle^2 \leqslant 1$$

and therefore, using (2.1) for $y = z = f_j$, $\sum \alpha_j^2 \leqslant n$. By the AM/GM inequality, this implies that $\left(\prod \alpha_j^2\right)^{1/n} \leqslant \frac{1}{n}\sum \alpha_j^2 \leqslant 1$. Since $\mathrm{vol}(\mathcal{E}) = \mathrm{vol}(B_2^n)\cdot\prod \alpha_j$, we conclude that $\mathrm{vol}(\mathcal{E}) \leqslant \mathrm{vol}(B_2^n)$.

Conversely, suppose that K is in John position. If $\frac{\mathrm{Id}}{n}$ does not belong to the convex set $\mathrm{conv}\{|x\rangle\langle x| \ : \ x \in S^{n-1}\cap\partial K\}$, then by the Hahn–Banach theorem there exists a linear form φ on M_n^{sa} such that $\varphi(\mathrm{Id}/n) < \varphi(|x\rangle\langle x|)$ for every $x \in \partial K \cap S^{n-1}$. Since M_n^{sa} is a Euclidean space for the inner product $(A,B) \mapsto \mathrm{Tr}(AB)$, the map φ has the form $\varphi(A) = \mathrm{Tr}(AH)$ for some H in M_n^{sa}. The hypothesis becomes $\frac{1}{n}\mathrm{Tr}(H) < \mathrm{Tr}(H|x\rangle\langle x|) = \langle x, Hx\rangle$ for every $x \in \partial K\cap S^{n-1}$. Finally, we may assume that $\mathrm{Tr}\, H = 0$ if we replace H by $H' = H - \frac{1}{n}\mathrm{Tr}\, H$. For $\delta > 0$ small enough, consider the ellipsoid

$$\mathcal{E}_\delta = \{x \in \mathbf{R}^n \ : \ \langle x, (\mathrm{Id} + \delta H)x\rangle \leqslant 1\}.$$

We claim that $\mathcal{E}_\delta \subset K$ for δ small enough. To check this, we compare the norms $\|\cdot\|_K$ with $\|\cdot\|_{\mathcal{E}_\delta}$. The latter can be computed as

$$\|x\|_{\mathcal{E}_\delta} = \inf\{t \geqslant 0 \ : \ x \in t\mathcal{E}_\delta\} = \sqrt{\langle x, (\mathrm{Id} + \delta H)x\rangle}.$$

It follows that

$$\|x\|_{\mathcal{E}_\delta}^2 - \|x\|_K^2 = \underbrace{\left(|x|^2 - \|x\|_K^2\right)}_{f(x)} + \delta\underbrace{\langle Hx, x\rangle}_{g(x)}.$$

The continuous functions f and g satisfy the following properties: $f \geqslant 0$ on S^{n-1} (since $B_2^n \subset K$), and $g > 0$ on the set $\{f = 0\}$. A little topological argument (check!) using the compactness of S^{n-1} implies that $f + \delta g > 0$ on S^{n-1} for δ small enough. It follows that there is $\varepsilon > 0$ such that $(1+\varepsilon)\mathcal{E}_\delta \subset K$.

Let (μ_j) be the eigenvalues of $\mathrm{Id} + \delta H$. We have $\sum \mu_j = n + \delta\,\mathrm{Tr}(H) = n$ and $\frac{\mathrm{vol}(\mathcal{E}_\delta)}{\mathrm{vol}(B_2^n)} = \prod \mu_j^{-1/2}$ (check!). By the AM/GM inequality, we have $(\prod \mu_j)^{1/n} \leqslant \frac{1}{n}\sum \mu_j = 1$ and therefore $\mathrm{vol}(\mathcal{E}_\delta) \geqslant \mathrm{vol}(B_2^n)$, so that $\mathrm{vol}((1+\varepsilon)\mathcal{E}_\delta) > \mathrm{vol}(B_2^n)$. This contradicts our hypothesis. □

2.3 Some distance estimates

Here are two corollaries of John's theorem.

Corollary 15. *For every symmetric convex body $K \subset \mathbf{R}^n$, we have $d_{BM}(K, B_2^n) \leqslant \sqrt{n}$.*

Corollary 16. *For every symmetric convex bodes $K, L \subset \mathbf{R}^n$, we have $d_{BM}(K, L) \leqslant n$.*

Proof of Corollary 15. We show that $\mathcal{E}_J(K) \subset K \subset \sqrt{n}\mathcal{E}_J(K)$. Since the problem is linearly invariant, we may assume that $\mathcal{E}_J(K) = B_2^n$. By John's theorem, there are contact points (x_i) in $\partial K \cap S^{n-1}$ and a convex combination (λ_i) such that $\frac{\mathrm{Id}}{n} = \sum \lambda_i |x_i\rangle\langle x_i|$. For every $x \in K$, we have $\langle x, x_i\rangle \leqslant 1$ (check!) and therefore

$$|x|^2 = \langle x, x\rangle = n\sum_i \lambda_i \langle x, x_i\rangle\langle x, x_i\rangle \leqslant n.$$

This proves the inclusion $K \subset \sqrt{n}B_2^n$ □

Theorem 17. *The metric space* (BM_n, d_{BM}) *is compact.*

Proof. Let (K_k) a sequence in BM_n. We may choose K_k such that $B_2^n \subset K_k \subset \sqrt{n}B_2^n$ for every k. Let $\|\cdot\|_k$ be the norm associated to K_kn which satisfies $\frac{1}{\sqrt{n}}|\cdot| \leqslant \|\cdot\|_k \leqslant |\cdot|$. For every k, the function $\|\cdot\|_k$ is 1-Lipschitz on S^{n-1} (check!). By Ascoli's theorem, there is a subsequence $\|\cdot\|_{\sigma(k)}$ which converges uniformly to a limit function $\|\cdot\|_{\text{lim}}$. We extend $\|\cdot\|_{\text{lim}}$ to a norm on $\mathbf{R}^n$ by setting

$$\|x\|_{\text{lim}} = |x| \cdot \left\| \frac{x}{|x|} \right\|_{\text{lim}} = \lim_{k\to\infty} \|x\|_{\sigma(k)}.$$

It is checked (check!) that uniform convergence on the sphere translates into the fact that $(K_\sigma(k))$ converges to K_{lim} in BM_n. $\square$

2.4 Distance between usual spaces

What is the value of $d_{BM}(K, L)$ as $n \to \infty$, when $K, L \in \{B_1^n, B_2^n, B_\infty^n\}$? We first start with the easiest case.

Proposition 18. *For every* n*, we have* $d_{BM}(B_1^n, B_2^n) = d_{BM}(B_\infty^n, B_2^n) = \sqrt{n}$.

Proof. The first equality is immediate by polarity. The $\leqslant$ inequality in the second one follows from Corollary 15. For the $\geqslant$ inequality, assume that $\alpha B_2^n \subset T(B_1^n) \subset B_2^n$ for some $T \in \mathsf{GL}_n(\mathbf{R})$. Denote $x_i = T(e_i)$ and observe that

$$T(B_1^n) = T(\text{conv}\{\pm e_i\}) = \text{conv}\{\pm x_i\}.$$

Moreover, for every $\varepsilon \in \{-1,1\}^n$, we have $\|\sum \varepsilon_i x_i\|_{T(B_1^n)} = n$. Consider now ε to be uniformly distributed on $\{-1,1\}^n$. By induction on n, using the parallelogram identity, we show (check!) that

$$\mathbf{E}_\varepsilon \left| \sum_{i=1}^n \varepsilon_i x_i \right|^2 = \sum_{i=1}^n |x_i|^2 \leqslant n.$$

The inclusion $\alpha B_2^n \subset T(B_1^n)$ implies $\|\cdot\|_{T(B_1^n)} \leqslant \alpha^{-1}|\cdot|$, and therefore

$$n^2 = \mathbf{E}_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_{T(B_1^n)}^2 \leqslant \alpha^{-2}\, \mathbf{E} \left| \sum_{i=1}^n \varepsilon_i x_i \right|^2 \leqslant \alpha^{-2} n.$$

We conclude that $\alpha^{-2} \geqslant n$, or $\alpha \leqslant \frac{1}{\sqrt{n}}$. $\square$

The case of estimating $d_{BM}(B_1^n, B_\infty^n)$ is more tricky. The upper bound $d_{BM}(B_1^n, B_\infty^n) \leqslant n$ is certainly not sharp for $n = 2$ since $d_{BM}(B_1^2, B_\infty^2) = 1$. The correct order of magnitude is $\sqrt{n}$.

Theorem 19. *For every n, we have*

$$c\sqrt{n} \leqslant d_{BM}(B_1^n, B_\infty^n) \leqslant C\sqrt{n}$$

where c, C are absolute constants (the proof gives $c = 1/\sqrt{2}$ and $C = 1 + \sqrt{2}$).

We first show the lower bound. We construct a sequence of matrices (W_k) as follows: W_k is a $2^k \times 2^k$ matrix, given by $W_0 = [1]$ and

$$W_{k+1} = \begin{pmatrix} W_k & W_k \\ W_k & -W_k \end{pmatrix}.$$

By construction, W_k is self-adjoint, with entries in $\{-1, 1\}$. Moreover it can be checked by induction on k (check!) that the columns of W_k are orthogonal, so that the matrix $2^{-k/2}W_k$ is orthogonal. We have $W_k(B_1^{2^k}) \subset B_\infty^{2^k}$ (since the entries of W_k are bounded by 1) and

$$W_k(B_1^{2^k}) \supset W_k(2^{-k/2}B_2^{2^k}) = B_2^{2^k} \supset 2^{-k/2}B_\infty^{2^k}.$$

This shows that $d_{BM}(B_1^n, B_\infty^n) \leqslant \sqrt{n}$ whenever n is a power of 2.

For the general case, we define by induction a $n \times n$ matrix A_n by

$$A_n = \begin{pmatrix} W_k & 0 \\ 0 & A_m \end{pmatrix}$$

where $n = 2^k + m$, $m < 2^k$. The matrix A_n has entries in $\{0, -1, 1\}$ and therefore $A_n(B_1^n) \subset B_\infty^n$. Let us check that

$$A_n(B_1^n) \supset \frac{1}{C\sqrt{n}} B_\infty^n$$

by induction on n. This is equivalent to $A_n^{-1} \subset C\sqrt{n}B_1^n$. We have

$$A_n^{-1} = \begin{pmatrix} W_k^{-1} & 0 \\ 0 & A_m^{-1} \end{pmatrix}$$

and therefore

$$\sup_{x \in B_\infty^n} \|A_n^{-1}x\|_1 = \sup_{x_1 \in B_\infty^{2^k}} \|W_k^{-1}x_1\|_1 + \sup_{x_2 \in B_\infty^m} \|A_m^{-1}x_2\|_1 \leqslant 2^{k/2} + C\sqrt{m}$$

where the last inequality uses the induction hypothesis. The induction is complete provided $2^{k/2} + C\sqrt{m} \leqslant C\sqrt{2^k + m}$ for every $m < 2^k$. One can verify (check!) that this holds for the choice $C = 1 + \sqrt{2}$.

The lower bound combines two classical inequalities which we now introduce. By a random sign we mean a random variable uniformly distributed on $\{-1, 1\}$. Khintchine inequalities says that the L^p norm are independent on the vector space spanned by an infinite sequence of independent random signs.

Proposition 20 (Khintchine inequalities)**.** *For every $p \in [1,2]$, there is a constant $A_p > 0$ and for every $p \in [2,\infty)$ there is a constant $B_p < \infty$ such that the following holds: if (ε_n) is a sequence of i.i.d. random signs, then for every n and every real numbers $a_1, \dots, a_n$, we have*

$$\forall p \in [2,\infty), \quad \left(\sum a_i^2\right)^{1/2} \leqslant \left(\mathbf{E}\left|\sum_{i=1}^n \varepsilon_i a_i\right|^p\right)^{1/p} \leqslant B_p \left(\sum a_i^2\right)^{1/2},$$

$$\forall p \in [1,2), \quad A_p \left(\sum a_i^2\right)^{1/2} \leqslant \left(\mathbf{E}\left|\sum_{i=1}^n \varepsilon_i a_i\right|^p\right)^{1/p} \leqslant \left(\sum a_i^2\right)^{1/2}.$$

Note that $A_2 = B_2 = 1$ by the parallelogram identity. It also holds that $B_p = O(\sqrt{p})$ as $p \to \infty$ (see Exercise 2.8) and that $A_1 = 1/\sqrt{2}$ (see Exercise 2.9).

We also need the following (check!)

Proposition 21 (Hadamard's inequality)**.** *Let $A \in \mathsf{M}_n$, and $v_1, \dots, v_n$ the columns of A. Then*

$$|\det A| \leqslant \prod_{i=1}^n |v_i|$$

and therefore

$$|\det A|^{1/n} \leqslant \frac{1}{n} \sum |v_i|.$$

We now prove that $d_{BM}(B_1^n, B_\infty^n) \geqslant c\sqrt{n}$. It suffices to show that is $A \in \mathsf{GL}_n(\mathbf{R})$ satisfies

$$\alpha^{-1} B_\infty^n \subset A(B_1^n) \subset B_\infty^n$$

then $\alpha \geqslant c\sqrt{n}$. Since $A(B_1^2) \subset \sqrt{n} B_2^n$, the columns of the matrix have a Euclidean norm at most $\sqrt{n}$, and therefore $|\det(A)| \leqslant n^{n/2}$ by Hadamard's inequality. On the other hand, since $A^{-1}(B_\infty^n) \subset \alpha B_1^n$, we have

$$\sup_{x \in B_\infty^n} \|A^{-1}x\|_1 \leqslant \alpha$$

If we denote by $L_1, \cdots, L_n$ the lines of the matrix A^{-1}, then

$$\begin{aligned} \alpha &\geqslant \sup_{\varepsilon \in \{-1,1\}^n} \sum_{i=1}^n |\langle L_i, \varepsilon \rangle| \\ &\geqslant \mathbf{E}_\varepsilon \sum_{i=1}^n |\langle L_i, \varepsilon \rangle| \\ &\geqslant \frac{1}{\sqrt{2}} \sum_{i=1}^n |L_i| \\ &\geqslant \frac{1}{\sqrt{2}} n |\det A^{-1}|^{1/n}, \end{aligned}$$

where we used Khintchine (with the value $A_1 = 1/\sqrt{2}$) and Hadamard inequalities. Since $|\det(A)| \leqslant n^{n/2}$, we have $|\det(A^{-1})|^{1/n} \geqslant n^{-1/2}$. It follows that $\alpha \geqslant \sqrt{n/2}$, as claimed.

Chapter 3

Concentration of measure

3.1 Volume of spherical caps

We denote by σ the uniform probability measure on the sphere S^{n-1}. It can be defined as follows: for a Borel set $A \subset S^{n-1}$, define

$$\sigma(A) = \frac{\mathrm{vol}_n(\{ta \ : \ t \in [0,1], a \in A\})}{\mathrm{vol}_n(B_2^n)}.$$

The measure σ is invariant under rotations: for any Borel set $A \subset S^{n-1}$ and $O \in \mathsf{O}(n)$, we have $\sigma(A) = \sigma(O(A))$. The measure σ is the unique Borel probability measure on S^{n-1} with this property (check!).

The sphere S^{n-1} can be equipped with two natural distances:

- the Euclidean distance $d(x,y) = |x-y|$, induced from the Euclidean norm on $\mathbf{R}^n$,
- the geodesic distance g, related to the Euclidean distance by the formula

$$|x-y| = 2\sin\left(\frac{g(x,y}{2}\right).$$

Since both distance are in one-to-one correspondence, statement about one distance have immediate translations into the other one. Moreover, they are related by the inequalities

$$\frac{2}{\pi}g(x,y) \leqslant |x-y| \leqslant g(x,y).$$

Given $x \in S^{n-1}$ and $\theta \in [0,\pi]$, we denote by

$$C(x,\theta) = \{y \in S^{n-1} \ : \ g(x,y) \leqslant \theta\}$$

the *spherical cap* with center x and angle θ. It follows from the rotation invariance that

$$V_n(\theta) \coloneqq \sigma(C(x,\theta))$$

does not depend on $x \in S^{n-1}$. We note the simple formulas $V_n(\frac{\pi}{2}) = \frac{1}{2}$ and $V_n(\pi - \theta) = 1 - V_n(\theta)$. One can also prove the analytic formula (check!)

$$V_n(\theta) = \frac{\int_0^\theta \sin^{n-2} t \,\mathrm{d}t}{\int_0^\pi \sin^{n-2} t \,\mathrm{d}t},$$

for which one can derive (check!) the fact that, for fixed $\theta \in [0, \pi/2]$,

$$\lim_{n\to\infty} V_n(\theta)^{1/n} = \sin\theta. \tag{3.1}$$

This is a important phenomenon that plays a fundamental role: **the proportion of the sphere covered by a cap with a fixed angle tends to 0 exponentially fast in large dimensions.**

Proposition 22. *For every $t \in [0, \pi/2]$, we have*

$$V(t) \leqslant \frac{1}{2} \sin^{n-1} t.$$

The proof uses the following fact (check!): if K, L are convex bodies such that $K \subset L$, then $a(K) \leqslant a(L)$, where $a(\,\cdot\,)$ is the surface area, defined in (1.5).

Sketch of proof. The surface area covered a cap of angle t (which equals $a(B_2^n)V_n(t)$) is less that the surface area covered by a half-sphere of radius $\sin(t)$ (which equals $\frac{1}{2}a(\sin(t)B_2^n) = \sin^{n-1} t a(B_2^n)$), as a consequence of the above fact (draw a picture). The result follows. □

As a corollary, we can see check that all the measure in a high-dimensional sphere is located close to an equator. For $\varepsilon \in (0, \pi/2)$, consider the set

$$A = \{(x_1, \dots, x_n) \in S^{n-1} \ : \ |x_n| \leqslant \sin\varepsilon\}$$

which is the ε-neighbourhood of an equator in geodesic distance. We have

$$\sigma(A) = 1 - \sigma(S^{n-1} \setminus A) = 1 - 2V_n(\pi/2 - \varepsilon) \geqslant 1 - \cos(\varepsilon)^{n-1}$$

using Proposition 22. If we combine this with the elementary inequality $\cos(t) \leqslant \exp(-t^2/2)$ (check!), we get $\sigma(A) \geqslant 1 - \exp(-(n-1)\varepsilon^2/2)$. It can also be proved, and we will use it (without proof) since it gives nicer formulas, that for $n \geqslant 2$ we have

$$V_n(\pi/2 - \varepsilon) \leqslant \exp(-n\varepsilon^2/2). \tag{3.2}$$

3.2 Covering and packing

Let (K,d) be a compact metric space. We denote by $B(x,\varepsilon)$ the closed ball centered at $x \in K$ and with radius $\varepsilon > 0$.

- We say that a finite subset $\mathcal{N} \subset K$ is an ε-net if $K = \bigcup_{x \in \mathcal{N}} B(x,\varepsilon)$. Equivalently, this means that for every $y \in K$, there is $x \in \mathcal{N}$ such that $d(x,y) \leqslant \varepsilon$. Nets exists by compactness. We denote by $N(K,\varepsilon)$ (or $N(K,d,\varepsilon)$) the smallest cardinality of an ε-net.
- We say that a finite subset $\mathcal{P} \subset K$ is ε-separated if for every distinct $x,y \in \mathcal{P}$ we have $d(x,y) > \varepsilon$. We denote by $P(K,\varepsilon)$ (or $P(K,d,\varepsilon)$) the largest cardinality of an ε-separated set.

Two simple but important inequalities are given by

$$P(K,2\varepsilon) \leqslant N(K,\varepsilon) \leqslant P(K,\varepsilon). \tag{3.3}$$

To prove the left inequality, note that if $\mathcal{P}$ is a 2ε-separated set and if $\mathcal{N}$ is an ε-net, the map which sends $y \in \mathcal{P}$ to a $x \in \mathcal{N}$ such that $d(x,y) \leqslant \varepsilon$ is injective, and therefore $\operatorname{card}(\mathcal{P}) \leqslant \operatorname{card}(\mathcal{N})$. For the right inequality, simply notice that a maximal ε-separated set is an ε-net.

The following lemma will be extremely useful

Lemma 23. *For every $\varepsilon \in (0,1)$, we have*

$$N(S^{n-1}, |\cdot|, \varepsilon) \leqslant \left(1 + \frac{2}{\varepsilon}\right)^n \leqslant \left(\frac{3}{\varepsilon}\right)^n$$

Proof. Let $\{x_i\}_{i \in I}$ be maximal ε-separated set in S^{n-1}. Then the balls (in $\mathbf{R}^n$) with centered x_i and radius $\varepsilon/2$ are disjoint, and are all included inside $(1+\varepsilon/2)B_2^n$. Therefore,

$$\operatorname{card}(I) \operatorname{vol}\left(\frac{\varepsilon}{2} B_2^n\right) \leqslant \operatorname{vol}\left(\bigcup_{i \in I} B(x_i,\varepsilon)\right) \leqslant \operatorname{vol}\left(\left(1 + \frac{\varepsilon}{2} B_2^n\right)\right),$$

and the result follows. □

We now discuss more finely, at ε fixed, how fast the quantities $N(S^{n-1},\varepsilon)$ and $P(S^{n-1},\varepsilon)$ grow. It turns out to be more convenient to use the geodesic distance. We start with the inequalities (check!)

$$\frac{1}{V_n(\varepsilon)} \leqslant N(S^{n-1},g,\varepsilon) \leqslant P(S^{n-1},g,\varepsilon) \leqslant \frac{1}{V_n(\varepsilon/2)}.$$

Proposition 24. *For any $\varepsilon \in (0, \pi/2)$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log N(S^{n-1}, g, \varepsilon) = -\log \sin \varepsilon$$

Since $N(S^{n-1}, g, \varepsilon) \geqslant V_n(\varepsilon)^{-1}$, the lower bound follows from (3.1). For the upper bound, we prove the following estimate: if $\varepsilon = \varepsilon_1 + \varepsilon_2$ with $0 < \varepsilon_1 < \varepsilon_2$, then

$$N(S^{n-1}, g, \varepsilon) \leqslant \left\lceil \frac{1}{V_n(\varepsilon_1)} \log \left(\frac{V_n(\varepsilon_1)}{V_n(\varepsilon_2)} \right) \right\rceil + \frac{1}{V_n(\varepsilon_1)}. \tag{3.4}$$

Using (3.4), one can prove (check!) that $\limsup \frac{1}{n} \log N(S^{n-1}, g, \varepsilon) \leqslant -\log \sin \varepsilon_1$ for every $\varepsilon_1 < \varepsilon$.

Proof of (3.4). We use a random covering argument due to Rogers (1957). Fix $N = \left\lceil \frac{1}{V_n(\varepsilon_1)} \log \left(\frac{V_n(\varepsilon_1)}{V_n(\varepsilon_2)} \right) \right\rceil$ and let $(x_i)_{1 \leqslant i \leqslant N}$ be i.i.d. random points on S^{n-1} distributed according to σ. Consider the set

$$A = \bigcup_{i=1}^{N} C(x_i, \varepsilon_1).$$

We compute, using Fubini theorem and the fact that $x \in C(x_i, \varepsilon) \iff x_i \in C(x, \varepsilon)$

$$\mathbf{E}\, \sigma(S^{n-1} \setminus A) = (1 - V_n(\varepsilon_1))^N \leqslant \exp(-N V_n(\varepsilon_1)) \leqslant \frac{V_n(\varepsilon_2)}{V_n(\varepsilon_1)}.$$

In particular, there exist $(x_1, \ldots, x_N)$ such that $\sigma(S^{n-1} \setminus A) \leqslant \frac{V_n(\varepsilon_2)}{V_n(\varepsilon_1)}$. Consider now $\{C(y_j, \varepsilon_2) \ : \ 1 \leqslant j \leqslant M\}$ to be a maximal family of disjoint caps of angle ε_2 contained in $S^{n-1} \setminus A$. Using disjointedness, we obtain $M V_n(\varepsilon_2) \leqslant \sigma(S^{n-1} \setminus A)$ and therefore $M \leqslant \frac{1}{V_n(\varepsilon_1)}$. On the other hand, by maximality, we have

$$S^{n-1} \subset \bigcup_{i=1}^{N} C(x_i, \varepsilon_1 + \varepsilon_2) \cup \bigcup_{j=1}^{M} C(y_j, 2\varepsilon_2)$$

showing (using that $2\varepsilon_2 \leqslant \varepsilon$) that $N(S^{n-1}, g, \varepsilon) \leqslant N + M$. □

In contrast with the case of covering, we have a poor understanding of optimal packing in high-dimensional spheres. For example, for fixed ε, the value of

$$\lim_{n\to\infty} \frac{1}{n} \log P(S^{n-1}, g, \varepsilon)$$

is not known (even the existence of the limit is not clear). We may conjecture that the value equals $-\log \sin(\varepsilon)$ as well. This would mean that one cannot substantially beat the greedy algorithm to produce packings.

3.3 Isoperimetric inequality on S^{n-1}

Exactly as in the case of $\mathbf{R}^n$, we have an isoperimetric inequality on the sphere.

Theorem 25. *Let $A \subset S^{n-1}$ be a closed set, and let C be a spherical cap such that $\sigma(A) = \sigma(C)$. Then for every $\varepsilon > 0$, we have $\sigma(A_\varepsilon) \geqslant \sigma(C_\varepsilon)$, where*

$$X_\varepsilon = \{x \in S^{n-1} \ : \ \exists y \in X \ : \ g(x,y) \leqslant \varepsilon\}.$$

This is harder to prove than the $\mathbf{R}^n$ version because it cannot be derived from the Brunn–Minkowski inequality. One proof goes as follows: one can define a spherical version of the Steiner symmetrization, and then adapt the argument we used in the proof of the Santaló inequality.

Corollary 26. *Let $A \subset S^{n-1}$ be a closed set with $\sigma(A) = \frac{1}{2}$. Then*

$$\sigma(A_\varepsilon) \geqslant 1 - \frac{1}{2}\exp(-n\varepsilon^2/2).$$

Proof. If C is a half-sphere, we have $\sigma(A_\varepsilon) \geqslant \sigma(C_\varepsilon) = V_n(\pi/2 + \varepsilon) = 1 - V_n(\pi/2 - \varepsilon)$ and we can use the formula (3.2) □

It is possible to derive from the Brunn–Minkowski inequality a variant of Corollary 26 with worse constants.

Theorem 27. *Let $A, B \subset S^{n-1}$ be closed sets such that $g(x,y) \geqslant \varepsilon$ for every $x \in A$, $y \in B$. Then we have*

$$\sigma(A)\sigma(B) \leqslant \exp(-n\varepsilon^2/4).$$

In particular, when $\sigma(A) = 1/2$, we get $\sigma(A_\varepsilon) \geqslant 1 - 2\exp(-n\varepsilon^2/4)$.

Proof. We define $\tilde{A} = \{tx \ : \ t \in [0,1], \ x \in A\}$, $\tilde{B} = \{tx \ : \ t \in [0,1], \ x \in B\}$ and note that (this is how we defined σ) $\mathrm{vol}(\tilde{A}) = \sigma(A)\,\mathrm{vol}(B_2^n)$ and $\mathrm{vol}(\tilde{B}) = \sigma(B)\,\mathrm{vol}(B_2^n)$. It follows then from the Brunn–Minkowski inequality that

$$\sqrt{\sigma(A)\sigma(B)}\,\mathrm{vol}(B_2^n) = \sqrt{\mathrm{vol}(\tilde{A})\,\mathrm{vol}(\tilde{B})} \leqslant \mathrm{vol}\left(\frac{\tilde{A}+\tilde{B}}{2}\right).$$

We now claim that $\frac{\tilde{A}+\tilde{B}}{2} \subset \cos(\varepsilon/2)B_2^n$. This is because the maximum of $\left|\frac{sx+ty}{2}\right|$ under the constraints $s,t \in [0,1]$ and $g(x,y) \geqslant \varepsilon$ is achieved for $s = t = 1$ and $g(x,y) = \varepsilon$ (check!). We have therefore

$$\sqrt{\sigma(A)\sigma(B)}\,\mathrm{vol}(B_2^n) \leqslant \mathrm{vol}(\cos(\varepsilon/2)B_2^n)$$

and the result follows using the inequality $\cos t \leqslant \exp(-t^2/2)$. □

In the following, we are going to use Corollary 26 even if we only proved the weaker version from Theorem 27.

A very important corollary is the following statement, sometimes known as Lévy's lemma.

Theorem 28 (Lévy's lemma)**.** *Let $f : (S^{n-1}, g) \to \mathbf{R}$ a 1-Lipschitz function, and M_f a median for f (i.e. a number which satisfies $\sigma(f \geqslant M_f) \geqslant \frac{1}{2}$, $\sigma(f \geqslant M_f) \leqslant \frac{1}{2}$. Then, for every $t > 0$ we have*

$$\sigma(f \geqslant M_f + t) \leqslant \frac{1}{2} \exp(-nt^2/2)$$

and therefore

$$\sigma(|f - M_f| \geqslant t) \leqslant \exp(-nt^2/2)$$

Remark. *1. In this context there is a unique median.*

2. If $f : (S^{n-1}, |\cdot|)$ is 1-Lipschitz, then it is also 1-Lipschitz for the geodesic distance, and the result applies.

3. If f is L-Lipschitz, Lévy's lemma applied to f/L gives $\sigma(|f - M_f| \geqslant t) \leqslant \exp(-nt^2/2L^2)$.

Proof. Let $A = \{x \in S^{n-1} \ : \ f(x) \leqslant M_f\}$. We have $\sigma(A) \geqslant \frac{1}{2}$. Since f is 1-Lipschitz, we have $f(x) \leqslant M_f + t$ for every $x \in A_t$, and therefore

$$\{f \geqslant M_f + t\} \subset S^{n-1} \setminus A_t\}.$$

It follows from Corollary 26 that

$$\sigma(\{f > M_f + t\}) \leqslant 1 - \sigma(A_t) \leqslant \frac{1}{2} \exp(-nt^2/2).$$

The second part is obtain by applying the result to $-f$:

$$\sigma(\{f < M_f - t\}) = \sigma(\{-f > M_{-f} + t\}) \leqslant \frac{1}{2} \exp(-nt^2/2).$$ □

It sometimes easier to compute the expectation $\mathbf{E}\, f$ rather than the median M_f. However, concentration of measure implies that $\mathbf{E}\, f$ and M_f are close to each other, and therefore a version of Lévy's lemma for expectation can be derived formally from Theorem 28 (check!).

Corollary 29. *Let $f : (S^{n-1}, g) \to \mathbf{R}$ a 1-Lipschitz function. Then, for every $t > 0$ we have*

$$\sigma(|f - \mathbf{E}[f]| \geqslant t) \leqslant C \ exp(-cnt^2)$$

for some absolute constants $C < \infty$ and $c > 0$

3.4 Gaussian concentration of measure

Let $(G_i)_{1\leqslant i\leqslant n}$ be i.i.d. $N(0,1)$ random variables, and $f:(\mathbf{R}^n,|\cdot|)\to\mathbf{R}$ a 1-Lipschitz function. Can we say something about the concentration of the random variable $X = f(G_1,\dots,G_n)$? Yes, and this turns out to be a corollary of the case of the sphere, thanks to the following phenomenon. We denote by γ_n the standard Gaussian distribution on $\mathbf{R}^n$, i.e. the law of $(G_1,\dots,G_n)$.

Theorem 30. *For $n\leqslant N$, identify $\mathbf{R}^n$ with a subspace of $\mathbf{R}^N$, and let $\pi_{N,n}:\sqrt{N}S^{N-1}\to\mathbf{R}^n$ be the orthogonal projection. Let $\mu_{N,n}$ be the image-measure under $\pi_{N,n}$ of the uniform probability measure on the sphere $\sqrt{N}S^{N-1}$. Then, for every n, as N to infinity, the sequence $(\mu_{N,n})_{N\geqslant n}$ converges in distribution towards γ_n.*

The uniform measure on the sphere $\sqrt{N}S^{N-1}$, which we denote σ_N, is understood as the image of σ under the map $x\mapsto\sqrt{N}x$.

Proof. For a Borel set $A\subset\mathbf{R}^n$, we have

$$\mu_{N,n}(A)=\sigma(\{x\in S^{N-1}\ :\ \pi_{N,n}(\sqrt{N}x)\in A\}).$$

Let $G=(G_1,\dots,G_N)$ a random vector with i.i.d. $N(0,1)$ entries. Since the distribution of G is invariant under rotation, the random vector $\frac{G}{|G|}$ is distributed according to the uniform measure on S^{N-1}. The measure $\mu_{N,n}$ is therefore the distribution of

$$\frac{\sqrt{N}}{(G_1^2+\cdots G_N^2)^{1/2}}(G_1,\dots,G_n).$$

By the law of large numbers, the prefactor $\frac{\sqrt{N}}{(G_1^2+\cdots G_N^2)^{1/2}}$ converges almost surely to 1, and the result follows. □

In turns out that a stronger notion of convergence holds: for any Borel set $A\subset\mathbf{R}^n$, we have

$$\lim_{N\to\infty}\mu_{N,n}(A)=\gamma_n(A). \tag{3.5}$$

Proving (3.5) for every Borel set is not so easy. When $\gamma_n(\partial A)=0$ (which is equivalent to $\mathrm{vol}(\partial A)=0$), the result follows from Portmanteau's theorem. This case will be sufficient for us (check! by adapting the following proof), as we will use (3.5) for sets of the form $A=B_\varepsilon$ (the ε-enlargement of B). Indeed, it can be checked (check! – use the Lebesgue differentiation theorem) that $\mathrm{vol}(\partial(B_\varepsilon))=0$ for every Borel set $B\subset\mathbf{R}^n$ and $\varepsilon>0$.

We now state the isoperimetric inequality for the Gaussian space $(\mathbf{R}^n,|\cdot|,\gamma_n)$

Corollary 31. *Let $A\subset\mathbf{R}^n$ be a Borel set, and H a half-space such that $\gamma_n(A)=\gamma_n(H)$. Then, for every $\varepsilon>0$, we have*

$$\gamma_n(A_\varepsilon)\geqslant\gamma_n(H_\varepsilon).$$

Equivalently, if we define $a \in [-\infty, +\infty]$ by the relation $\gamma_n(A) = \gamma_1((-\infty, a])$, we have $\gamma_n(A_\varepsilon) \geqslant \gamma_1((-\infty, a+\varepsilon])$.

Proof. If $\gamma_n(A) = 0$ or $\gamma_n(A) = 1$ the result is obvious. Otherwise, define $a \in \mathbf{R}$ by the relation $\gamma_n(A) = \gamma_1((-\infty, a])$. For every $b < a$, we have $\gamma_n(A) > \gamma_1((\infty, b])$. Since

$$\gamma_n(A) = \lim_{N\to\infty} \sigma_N(\pi_{N,n}^{-1}(A)) \quad \text{and} \quad \gamma_1((\infty, b]) = \lim_{N\to\infty} \sigma_N(\pi_{N,1}^{-1}((-\infty, b]),$$

we have $\sigma_N(\pi_{N,n}^{-1}(A)) \geqslant \sigma_N(\pi_{N,1}^{-1}((-\infty, b])$ for N large enough. Since the set $\pi_{N,1}^{-1}((-\infty, b])$ is a spherical cap in $\sqrt{N}S^{N-1}$, the spherical isoperimetric inequality implies that

$$\sigma_N(\pi_{N,n}^{-1}(A)_\varepsilon) \geqslant \sigma_N(\pi_{N,1}^{-1}((-\infty, b])_\varepsilon)$$

where ε-enlargements are taken with respect to the geodesic distance on $\sqrt{N}S^{N-1}$. We check that $\pi_{N,n}^{-1}(A)_\varepsilon \subset \pi_{N,n}^{-1}(A_\varepsilon)$. On the other hand, we have (check!)

$$\pi_{N,1}^{-1}((-\infty, b])_\varepsilon) = \pi_{N,1}^{-1}((-\infty, \varepsilon_N)$$

where the number ε_N is defined by the relations $\sin(\theta_N) = \frac{b}{\sqrt{N}}$ and $\sin(\theta_N + \frac{\varepsilon}{\sqrt{N}}) = \frac{b+\varepsilon_N}{\sqrt{N}}$. The numbers (ε_N) tend to ε as N tends to infinity (check!), and therefore, using (3.5) twice, we obtain

$$\gamma_n(A_\varepsilon) \geqslant \gamma_1((-\infty, b+\varepsilon)).$$

The last step is to take the supremum over $b < a$. □

As in the case of the sphere, we have (same proof, check!)

Corollary 32. *Let $F : \mathbf{R}^n \to \mathbf{R}$ be a 1-Lipschitz function with respect to the Euclidean distance, and $G_1, \cdots, G_n$ be i.i.d. $N(0,1)$ random variables. If M_X is the median of $X = F(G_1, \ldots, G_n)$, then for every $t > 0$,*

$$\mathbf{P}(X \geqslant M_X + t) \leqslant \mathbf{P}(G_1 \geqslant t).$$

Some sharp inequalities are know on the quantity

$$\mathbf{P}(G_1 \geqslant t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp(-x^2/2)\,\mathrm{d}x.$$

For example, one has the Komatsu inequalities for $x > 0$

$$\frac{2}{x + \sqrt{x^2+4}} \leqslant e^{x^2/2} \int_x^\infty e^{-t^2/2}\,\mathrm{d}t \leqslant \frac{2}{x + \sqrt{x^2+2}} \tag{3.6}$$

which give a sharp bound when $t \to \infty$. Another simple bound is the inequality (check!)

$$\mathbf{P}(G_1 \geqslant t) \leqslant \frac{1}{2}\exp(-t^2/2).$$

It follows that, in the context of Corollary 32, we have

$$\mathbf{P}(X \geqslant M_X + t) \leqslant \frac{1}{2}\exp(-t^2/2),$$

$$\mathbf{P}(|X - M_X| \geqslant t) \leqslant \exp(-t^2/2),$$

As an application of Gaussian concentration, we prove the Johnson–Lindenstrauss lemma. The context is the following: we have a finite set $A \subset \mathbf{R}^n$, and we search for a linear map $f : \mathbf{R}^n \to \mathbf{R}^k$, for $k \ll n$, which is almost an isometry when restricted to A, in the sense that for every $x, y \in A$, we have

$$(1-\varepsilon)|x-y| \leqslant |f(x) - f(y)| \leqslant (1+\varepsilon)|x-y|.$$

When $\varepsilon = 0$ the best possible is $k = \min(n, \mathrm{card}(A))$. Remarkably, for any $\varepsilon > 0$, this can be greatly improved to k of order $\log \mathrm{card}(A)$.

Theorem 33 (Johnson–Lindenstrauss lemma)**.** *Let $A \subset \mathbf{R}^n$, $m = \mathrm{card}(A)$ and $\varepsilon \in (0,1)$. If $k \geqslant C\log(m)/\varepsilon^2$, there is a linear map $f : \mathbf{R}^n \to \mathbf{R}^k$ such that for every $x, y \in A$,*

$$(1-\varepsilon)|x-y| \leqslant |f(x) - f(y)| \leqslant (1+\varepsilon)|x-y|.$$

Proof. Pick f at random! Let $B : \mathbf{R}^n to \mathbf{R}^k$ be a random linear map corresponding to a matrix $(b_{ij})_{1 \leqslant i \leqslant k, 1 \leqslant j \leqslant n}$ with i.i.d. $N(0,1)$ entries. The following remark is fundamental (check!): for every $u \in S^{n-1}$ the random vector Bu has distribution γ_k. Moreover, since the function $x \mapsto |x|$ is 1-Lipschitz, we have

$$\mathbf{P}\left(\Big||Bu| - M_k\Big| \geqslant t\right) \leqslant \exp(-t^2/2) \tag{3.7}$$

for every $u \in S^{n-1}$, where M_k is the random variable $X = \sqrt{G_1^2 + \cdot + G_k^2}$, with (G_i) i.i.d. $N(0,1)$. It can be checked (check!) that M_k is of order $\sqrt{k}$ (by concentration of measure, all the quantities M_X, $\mathbf{E}\,X$ and $(\mathbf{E}\,X^2)^{1/2} = \sqrt{k}$ differ my at most $O(1)$).

Define $f = \frac{1}{M_k}B$. Given $x \neq y$ in A, we apply (3.7) to the unit vector $u = \frac{x-y}{|x-y|}$ and $t = \varepsilon M_k$ to obtain

$$\mathbf{P}\left(\Big||f(x) - f(y)| - |x-y|\Big| \geqslant \varepsilon|x-y|\right) \leqslant \exp(-\varepsilon^2 M_k^2/2).$$

Therefore, by the union bound

$$\mathbf{P}\left(\exists x \neq y \in A \ : \ \left|\frac{|f(x)-f(y)|}{[x-y|} - 1\right| > \varepsilon\right) \leqslant \binom{m}{2}\exp(-\varepsilon^2 M_k^2/2).$$

The right-hand side is less that 1 (and therefore, there exists a f with the desired property) whenever $\varepsilon^2 M_k^2/2 \geqslant \log\binom{m}{2}$, which is satisfied provided $k \geqslant C\log(m)/\varepsilon^2$ since $M_k \sim \sqrt{k}$. □

Chapter 4

Dvoretzky's theorem

4.1 Background

We denote by $\ell_2^n = (\mathbf{R}^n, |\cdot|)$ the n-dimensional Euclidean space.

We start with the following question, which was asked by Grothendieck: is it true for every $n \in \mathbf{N}^*$ and $\varepsilon > 0$, every infinite-dimensional Banach space X contains an n-dimensional subspace Y such that $d_{BM}(Y, \ell_2^n) \leqslant 1 + \varepsilon$.

As a warm-up we show that the question has an easy positive answer for the special case of $X = L^p([0,1])$ (with $1 \leqslant p < \infty$). The idea is to construct on the probability space $([0,1], \mathrm{vol})$ an i.i.d. sequence of $N(0,1)$ random variables (G_n). (For example (check!), use the binary expansion of an element in $[0,1]$ to obtain an infinite sequence of i.i.d. Bernoulli(1/2) variables, which can be used to simulate any distribution). For any real numbers $a_1, \dots, a_n$, observe that $a_1 G_1 + \cdots + a_n G_n$ has distribution $N(0, |a|^2)$, and therefore

$$\left\| \sum_{i=1}^n a_i G_i \right\|_{L^p} = \alpha_p |a|$$

where α_p is the L^p-norm of a $N(0,1)$ random variable. This shows that the space $Y = \mathrm{span}(G_1, \dots, G_n) \subset L^p([0,1])$ is isometric to ℓ_2^n.

The general case is more involved. We are going to prove the following theorem, which implies a positive answer to Grothendieck's question (check!).

Theorem 34. *For every $\varepsilon > 0$, there is a constant $c(\varepsilon) > 0$ such that every n-dimensional normed space X admits a k-dimensional subspace E with $k = \lfloor c(\varepsilon) \log(n) \rfloor$ such that*

$$d_{BM}(E, \ell_2^k) \leqslant 1 + \varepsilon.$$

We can obtain an equivalent statement about symmetric convex bodies: for every $\varepsilon > 0$, there is a constant $c(\varepsilon) > 0$ such that, whenever $K \subset \mathbf{R}^n$ is a symmetric convex body, there is a subspace $E \subset \mathbf{R}^n$ with $k = \dim E = \lfloor c(\varepsilon) \log(n) \rfloor$ such that $d_{BM}(K \cap E, B_2^k) \leqslant 1 + \varepsilon$. The section $K \cap E$ is "almost ellipsoidal".

As an example, we work out the case of B_∞^n. We are looking for a linear map $A : \mathbf{R}^k \to \mathbf{R}^n$ such that

$$\frac{1}{1+\varepsilon}|x| \leqslant \|A(x)\|_\infty \leqslant |x|$$

for every $x \in \mathbf{R}^k$. The map A has the form

$$x \mapsto (\langle x, x_1 \rangle, \dots, \langle x, x_n \rangle)$$

for some vectors $x_1, \dots, x_n \in \mathbf{R}^k$. We have $|x_j| \leqslant 1$ and we may assume without generality that $|x_j| = 1$ (replace x_j by $\frac{x_j}{|x_j|}$). We have therefore, for every $x \in \mathbf{R}^k$,

$$\max_{1 \leqslant j \leqslant n} |\langle x, x_j \rangle| \geqslant \frac{1}{1+\varepsilon}|x|.$$

This is equivalent (check!) to the fact that $\operatorname{conv}\{\pm x_i\} \supset \frac{1}{1+\varepsilon} B_2^n$, and also equivalent (check!) to the fact that (x_j) is θ-net in (S^{k-1}, g), for $\cos\theta = \frac{1}{1+\varepsilon}$. From the estimates on the size of nets in the sphere, we know that such vectors $(x_j)_{1 \leqslant j \leqslant n}$ exist if $n \geqslant \exp(C(\varepsilon)k)$, and that this behaviour is sharp (up to the value of $C(\varepsilon)$). Therefore, the logarithmic dependence in Theorem 34 is optimal.

4.2 Haar measure

Any compact group (=a group which is also a compact topological space, such that the group operations $g \mapsto g^{-1}$ and $(g, h) \mapsto gh$ are continuous) carries a unique Haar probability measure

Theorem 35. *If G is a compact group, there exists a unique Borel probability measure μ_H (the Haar measure) which is invariant under left- and right- translations, i.e. such that for every $g \in \mathsf{G}$ and Borel set $A \subset \mathsf{G}$,*

$$\mu_H(g \cdot A) = \mu_H(A \cdot g) = \mu_H(A).$$

We are going work with the Haar measure on the group O_n. In this case the Haar measure can be described explicitly as follows. We give an algorithm to construct a random element $O \in \mathsf{O}_n$. We first choose a random vector $e_1 \in S^{n-1}$ according to σ. Then, we choose e_2 at random on the sphere $S^{n-1} \cap e_1^\perp$, which we identify with S^{n-1}, according to σ. We iterate this process and define by induction $(e_1, \dots, e_n)$ by choosing e_k according to the measure σ on the sphere $S^{n-1} \cap \{e_1, \dots, e_{k-1}\}^\perp$, identified with S^{n-k}. To define the last vector e_k, we choose with probability $\frac{1}{2}$ one of the two elements of $S^{n-1} \cap \{e_1, \dots, e_{n-1}\}^\perp$. All the choices are made independently. We then consider the matrix O whose columns are given by $(e_1, \dots, e_n)$. By construction O is an orthogonal matrix, and it can be checkde (using the fact that the measure σ is invariant under rotations) that the distribution of O is the Haar measure.

For $0 \leqslant k \leqslant n$, we denote by $\mathsf{G}_{n,k}$ the family of all k-dimensional subspaces of $\mathbf{R}^n$. The set $\mathsf{G}_{n,k}$ is called the *Grassmann manifold*. It can be equipped with a metric by the formula $d(E,F) = \|P_E - P_F\|_\infty$, where P_E is the orthonormal projection onto E.

The group $\mathsf{O}(n)$ acts transitively on $\mathsf{G}_{n,k}$ (in the following sense: for every E, $F \in \mathsf{G}_{n,k}$, there is $O \in \mathsf{O}(n)$ such that $O(E) = F$). Therefore, if $O \in \mathsf{O}(n)$ is Haar distributed, the distribution of $O(E)$ is the same for any $E \in \mathsf{G}_{n,k}$, and will be denoted by $\mu_{n,k}$. More concretely, one can define $\mu_{n,k}$ as the distribution of

$$\operatorname{span}\{x_1, \dots, x_k\}$$

where (x_i) are i.i.d. random points in S^{n-1} with distribution σ, or equivalently as the distribution of

$$\operatorname{span}\{G_1, \dots, G_k\}$$

where (G_i) are i.i.d. Gaussian vectors with distribution $N(0, \mathrm{Id})$.

An important remark is that while the set $\mathsf{G}_{n,k}$ can be defined without referring to a Euclidean structure, the measure $\mu_{n,k}$ does depend on the underlying Euclidean structure.

We now state a theorem about concentration of Lipschitz functions on a subspace.

Theorem 36. *Let $f : (S^{n-1}, |\cdot|) \to \mathbf{R}$ a 1-Lipschitz function, with mean $\mathbf{E}[f]$. Let $E \in G_{n,k}$ be a random subspace with distribution $\mu_{n,k}$, and $\varepsilon \in (0,1)$. If $k \leqslant c(\varepsilon)n$, then with high probability,*

$$\sup_{x \in S^{n-1} \cap E} |f(x) - \mathbf{E}[f]| \leqslant \varepsilon,$$

where $c(\varepsilon) = c\varepsilon^2 / \log(1/\varepsilon)$, $c > 0$ being an absolute constant.

The theorem above is true with $c(\varepsilon) = c\varepsilon^2$, but requires a proof more sophisticated than the union bound argument. In this theorem, "with high probability" should be understood as follows: the probability of the complement is smaller than $C(\varepsilon)\exp(-c(\varepsilon)n)$.

Proof. By Corollary 29, if $y \in S^{n-1}$ is chosen at random according to the distribution σ,

$$\mathbf{P}\left(|f(y) - \mathbf{E}[f]| > \varepsilon\right) \leqslant C\exp(-cn\varepsilon^2). \tag{4.1}$$

Pick arbitrarily $E_0 \in \mathsf{G}_{n,k}$, and let $\mathcal{N}$ be a ε-net in $(S^{n-1} \cap E_0, |\cdot|)$. Since $S^{n-1} \cap E_0$ can be identified with S^{k-1}, we may enforce using Lemma 23 that $\operatorname{card} \mathcal{N} \leqslant (3/\varepsilon)^k$. Consider a random $O \in \mathsf{O}(n)$ with distribution μ_H, so that $O(E_0)$ has distribution $\mu_{n,k}$. Since $f \circ O$ is 1-Lipschitz, we have

$$\sup_{x \in S^{n-1} \cap E_0} |f(Ox) - \mathbf{E}[f]| \leqslant \varepsilon + \sup_{x \in \mathcal{N}} |f(Ox) - \mathbf{E}[f]|$$

and therefore, by the union bound

$$\begin{aligned}\mathbf{P}\left(\sup_{x\in S^{n-1}\cap E_0}|f(Ox)-\mathbf{E}[f]|>2\varepsilon\right) &\leqslant \sum_{x\in\mathcal{N}}\mathbf{P}\left(|f(Ox)-\mathbf{E}[f]|>\varepsilon\right)\\ &\leqslant (3/\varepsilon)^k C\exp(-cn\varepsilon^2),\end{aligned}$$

where we used (4.1) and the fact that for every $x \in S^{n-1}$, the distribution of Ox is σ (check!). If we denote $p = (3/\varepsilon)^k C\exp(-cn\varepsilon^2)$, we see that $p < 1$ provided $k \leqslant c(\varepsilon)n$ for some $c(\varepsilon) = c\varepsilon^2/\log(1/\varepsilon)$. Moreover, up to changing the value of constants, this condition implies that $p \leqslant C\exp(-cn\varepsilon^2)$. This completes the proof. □

4.3 Proof of the Dvoretzky–Milman theorem

We are going to use the following fact.

Proposition 37. *Let $K \subset \mathbf{R}^n$ be a symmetric convex body such that $\mathcal{E}_J(K) = B_2^n$. Then,*

$$\int_{S^{n-1}}\|x\|_K \,\mathrm{d}\sigma(x) \geqslant c\sqrt{\frac{\log n}{n}}.$$

Proof of Theorem 34. Since the problem is invariant under linear images, we may assume that $\mathcal{E}_J(K) = B_2^n$. We repeat the argument used in the proof of Theorem 36. Fix $E_0 \in \mathsf{G}_{n,k}$, and let $\mathcal{N}_0$ be a θ-net in $(S^{n-1}\cap E_0, |\cdot|)$ with $\operatorname{card}\mathcal{N}_0 \leqslant (3/\theta)^k$. Take $O \in \mathsf{O}(n)$ at random with distribution μ_H, and let $E = O(E_0)$. Note that $\mathcal{N} := O(\mathcal{N}_0)$ is a θ-net in $(S^{n-1}\cap E, |\cdot|)$. Consider the function $f = \|\cdot\|_K$ on S^{n-1}, which is 1-Lipschitz since $B_2^n \subset K$ (check!). By arguing as in the proof of Theorem 36, we obtain

$$\mathbf{P}\left(\sup_{x\in\mathcal{N}_0}|f(Ox)-\mathbf{E}[f]|\geqslant\eta\right) \leqslant \underbrace{\left(\frac{3}{\theta}\right)^k C\exp(-cn\eta^2)}_{p}$$

We choose $\eta = \varepsilon\,\mathbf{E}[f]$ and conclude that with probability at least $1-p$,

$$\forall x\in\mathcal{N},\quad (1-\varepsilon)\,\mathbf{E}[f] \leqslant f(x) = \|x\|_K \leqslant (1+\varepsilon)\,\mathbf{E}[f].$$

We claim that event this implies

$$\forall x \in S^{n-1}\cap E,\quad \left(1-\varepsilon-\frac{\theta(1+\varepsilon)}{1-\theta}\right)\mathbf{E}[f] \leqslant \|x\|_K \leqslant \frac{1+\varepsilon}{1-\theta}\,\mathbf{E}[f]. \tag{4.2}$$

To see this, consider $A = \sup\{\|x\|_K \ :\ x \in S^{n-1}\cap E\}$. Given $x \in S^{n-1}\cap E$, there is $y \in \mathcal{N}$ with $|x-y| \leqslant \theta$. Therefore,

$$\|x\|_K \leqslant \|y\|_K + \|x-y\|_K = \|y\|_K + |x-y|\left\|\frac{x-y}{|x-y|}\right\|_K \leqslant \|y\|_K + \theta A.$$

Taking supremum over x gives the inequality $A \leqslant \sup_{y\in\mathcal{N}} \|y\|_K + \theta A$, and thus the upper bound in (4.2). For the lower bound, we argue similarly that

$$\|x\|_K \geqslant \|y\|_K - \|x-y\|_K \geqslant \|y\|_K - \theta A.$$

If we choose $\theta = \varepsilon$, then (4.2) implies that (with probability at least $1-p$) for every $x \in S^{n-1} \cap E$

$$(1-3\varepsilon)\,\mathbf{E}[f]|x| \leqslant \|x\|_K \leqslant (1+3\varepsilon)\,\mathbf{E}[f]|x|.$$

If $p<1$, we can conclude that $d_{BM}(K\cap E, \ell_2^k) \leqslant \frac{1+3\varepsilon}{1-3\varepsilon}$, as wanted. It remains to analyze when $p<1$. The condition $p<1$ is equivalent to $k\log(3/\varepsilon) < cn\varepsilon^2\,\mathbf{E}[f]^2$. By Proposition 37, we have $\mathbf{E}[f] \geqslant c\sqrt{\frac{\log n}{n}}$, and therefore the condition $p<1$ is satisfied whenever $k < c\log(n)\varepsilon^2/\log(1/\varepsilon)$ □

4.4 Basic estimates on Gaussian variables

It remains to prove Proposition 37. To do this, it is useful to replace integrals over S^{n-1} by Gaussian integration. Let $G=(g_1,\dots,g_n)$ a vector with i.i.d. $N(0,1)$ coordinates. Then, the random variables $|G|$ and $\frac{G}{|G|}$ are independent, and the latter is distributed according to σ. This can be seen as follows: consider $O \in \mathsf{O}(n)$ independent from G, and with distribution μ_H. Then OG and G have the same distribution, and therefore $\left(|G|, \frac{G}{|G|}\right)$ and $\left(|G|, O\left(\frac{G}{|G|}\right)\right)$ also have the same distribution. Since Ox has distribution σ for an arbitrary $x \in S^{n-1}$, the claim follows.

A consequence is the formula, for any norm

$$\int_{S^{n-1}} \|x\|\,\mathrm{d}\sigma(x) = \frac{1}{\mathbf{E}\,|G|}\,\mathbf{E}\,\|G\|. \tag{4.3}$$

Indeed, we have $\mathbf{E}\,\|G\| = \mathbf{E}\left\||G|\frac{G}{|G|}\right\| = \mathbf{E}\,|G|\cdot\mathbf{E}\left\|\frac{G}{|G|}\right\|$ by independence. It is useful to denote by κ_n the number $\mathbf{E}\,|G|$. Basic estimates are

$$\kappa_n \leqslant \left(\mathbf{E}\,|G|^2\right)^{1/2} = \sqrt{n},$$

$$\kappa_n \geqslant \frac{1}{\sqrt{n}}\,\mathbf{E}\,\|G\|_1 = \sqrt{n}\sqrt{2/\pi},$$

and one may check that $\kappa_n \sim \sqrt{n}$ as $n\to\infty$ (check!).

We now state an elementary lemma about Gaussian variables. Essentially, the function $\sqrt{2\log x}$ appears since it is the inverse of the function $\exp(x^2/2)$.

Lemma 38. *Let $g_1,\dots,g_n$ be $N(0,1)$ random variables. Then $\mathbf{E}\max(g_i) \leqslant \sqrt{2\log n}$. If moreover the (g_i) are independent, then $\mathbf{E}\max(g_i) \geqslant c\sqrt{\log n}$ for some $c>0$.*

Proof. For the first part, we use the formula (check!) $\mathbf{E}\exp(tg_i) = \exp(t^2/2)$ for $t \in \mathbf{R}$. For $\beta > 0$ to be chosen later, we have (using Jensen's inequality and the concavity of log)

$$\begin{aligned} \mathbf{E}\max(g_1, \dots, g_n) &\leqslant \mathbf{E}\,\frac{1}{\beta}\log\sum_{i=1}^n \exp(\beta g_i) \\ &\leqslant \frac{1}{\beta}\log \mathbf{E}\sum_{i=1}^n \exp(\beta g_i) \\ &= \frac{1}{\beta}\log(n\exp(\beta^2/2)) \\ &= \frac{\log n}{\beta} + \frac{\beta}{2} \end{aligned}$$

and the optimal value $\beta = \sqrt{2\log n}$ gives the result. For the second part, we may write

$$\begin{aligned} \mathbf{P}(\max g_i > \alpha) &= 1 - \mathbf{P}(\max g_i \leqslant \alpha) \\ &= 1 - \mathbf{P}(g_1 \leqslant \alpha)^n \\ &= 1 - (1 - \mathbf{P}(g_1 > \alpha))^n \\ &\geqslant 1 - \exp(-n\mathbf{P}(g_1 > \alpha)) \end{aligned}$$

We now choose α such that $\mathbf{P}(g_1 > \alpha) = \frac{1}{n}$. It can be checked (check! – use e.g. (3.6)) that $\alpha \geqslant c\sqrt{\log n}$. We have therefore $\mathbf{E}\max(g_i) \geqslant \alpha\mathbf{P}(\max(g_i) \geqslant \alpha) \geqslant \alpha(1 - 1/e)$. □

4.5 Proof of Proposition 37

We start with a lemma

Lemma 39. *Let $K \subset \mathbf{R}^n$ be a symmetric convex body with $\mathbf{E}_J(K) = B_2^n$. Then there exists an orthonormal basis (x_k) of $\mathbf{R}^n$ such that $\|x_k\|_K \geqslant \sqrt{k/n}$.*

Proof. We iterate the following fact: any subspace $F \subset \mathbf{R}^n$ contains a vector x such that $|x| = 1$ and $\|x\|_K \geqslant \sqrt{\dim(F)/n}$. This fact is enough to construct by induction an orthonormal basis $(x_1, \dots, x_n)$ with $\|x_1\|_K \geqslant 1$, $\|x_2\|_K \geqslant \sqrt{(n-1)/n}$, $\cdots$, $\|x_n\|_K \geqslant \sqrt{1/n}$ (apply the fact to the subspace $F = \operatorname{span}\{x_1, \dots, x_{k-1}\}$. Note that using the fact with $\dim F = 1$ gives a proof that $K \subset \sqrt{n}B_2^n$ (Corollary 15).

We now prove the fact. By John's theorem, there exist contact points $x_i \in S^{n-1} \cap \partial K$, and a convex combination (λ_i) such that

$$\frac{\mathrm{Id}}{n} = \sum_i \lambda_i |x_i\rangle\langle x_i|.$$

Also remember that $\langle x, x_i \rangle \leqslant 1$ for any $x \in K$ (cf. the proof of John's theorem), and therefore $\| \cdot \|_K \geqslant \langle \cdot, x_i \rangle$. If we denote by P_F the orthonormal projection onto F, the previous inequality yields $\frac{P_F}{n} = \sum_i \lambda_i |P_F x_i\rangle\langle P_F x_i|$. Take the trace, we obtain

$$\frac{\dim F}{n} = \sum_i \lambda_i |P_F x_i|^2$$

and therefore there exists an index i such that $|P_F x_i| \geqslant \sqrt{\dim(F)/n}$. The vector $x = \frac{P_F x_i}{|P_F x_i|}$ has the desired property: indeed $|x| = 1$ and

$$\|x\|_K \geqslant \langle x, x_i \rangle = \frac{[P_F x_i|^2}{|P_F x_i|} \geqslant \sqrt{\dim(F)/n}. \qquad \square$$

We can now complete the proof of Proposition 37. We have

$$\int_{S^{n-1}} \| \, x \, \|_K \, \mathrm{d}\sigma(x) = \frac{1}{\kappa_n} \, \mathbf{E} \, \|G\|_K$$

Where $G = (g_1, \ldots, g_n)$ is a $N(O, \mathrm{Id})$ random vector. By Lemma 39, applying an orthogonal transformation if necessary, we may reduce to the case when the canonical basis (e_i) satisfies $\|e_i\|_K \geqslant \sqrt{i/n}$. We now use the following trick: if $(\varepsilon_1, \ldots, \varepsilon_n)$ are random signs independent from G, than $(\varepsilon_1 g_1, \ldots, \varepsilon_n g_n)$ has the same distribution as G. Therefore,

$$\begin{aligned}
\mathbf{E} \, \|G\|_K &= \mathbf{E}_g \, \mathbf{E}_\varepsilon \, \|(\varepsilon_1 g_1, \ldots, \varepsilon_n g_n)\|_K \\
&\geqslant \mathbf{E}_g \max_i \|g_i e_i\|_K \\
&\geqslant \mathbf{E} \max_{1 \leqslant i \leqslant n} |g_i| \sqrt{i/n} \\
&\geqslant \frac{1}{\sqrt{2}} \, \mathbf{E} \max_{n/2 \leqslant i \leqslant n} |g_i|
\end{aligned}$$

and we conclude that $\mathbf{E} \, \|G\|_K \geqslant c\sqrt{\log n}$ by the second part of Lemma 38.

Chapter 5

Gluskin's theorem

5.1 Preliminaries: on the volume of polytopes

We define a *polytope* to be a convex body which is the convex hull of a finite set. Equivalently (check!), a polytope is a convex body which is the intersection of finitely many closed half-spaces.

Let P a polytope. If $P = \operatorname{conv} A$, for a subset $A \subset \mathbf{R}^n$ which is minimal with this property, the elements of A are called the *vertices* of P. If $P = \bigcap H_i$ for a family (H_i) of half-spaces which is minimal with this property, then the convex sets $H_i \cap \partial P$ are called the *facets* of P.

A *simplex* in $\mathbf{R}^n$ is polytope with $n+1$ vertices, or equivalently with $n+1$ facets. When $0 \in \operatorname{int}(P)$, there is a one-to-one correspondence between the vertices of P and the facets of P°.

Let $K \subset \mathbf{R}^n$ be a convex body. For $u \in \mathbf{R}^n$, define

$$w_K(u) = \sup_{x \in K} \langle u, x \rangle,$$

which for $|u| = 1$ is called the *width* of K in the direction u. The *width* of K is the average of mean width over directions

$$w(K) = \int_{S^{n-1}} w(K, u) \, \mathrm{d}\sigma(u).$$

The mean width gives an upper bound on the volume, which is often quite good. It is convenient to write it using the *volume radius* of a convex body K, defined as

$$\operatorname{vrad}(K) = \left(\frac{\operatorname{vol}(K)}{\operatorname{vol}(B_2^n)} \right)^{1/n}.$$

So far, we never computed the value of $\operatorname{vol}(B_2^n)$. It equals

$$\operatorname{vol}(B_2^n) = \frac{\pi^{n/2}}{\Gamma\left(1 + \frac{n}{2}\right)}$$

from which it can be derived than $\text{vol}(B_2^n)^{1/n} \sim \frac{\sqrt{2e\pi}}{\sqrt{n}}$ as n tends to infinity. Here is a simple, but instructive, way to obtain the correct order of $\text{vol}(B_2^n)^{1/n}$. We start from the inequalities $\frac{1}{\sqrt{n}}B_\infty^n \subset B_2^n \subset \sqrt{n}B_1^n$ to obtain

$$\left(\frac{2}{\sqrt{n}}\right)^n = \text{vol}\left(\frac{1}{\sqrt{n}}B_\infty^n\right) \leqslant \text{vol}(B_2^n) \leqslant \text{vol}(\sqrt{n}B_1^n) = \frac{2^n n^{n/2}}{n!} \leqslant \left(\frac{C}{\sqrt{n}}\right)^n. \tag{5.1}$$

To compute $\text{vol}(B_1^n)$, observe that B_1^n is the union of 2^n simplices congruent to $\text{conv}(e_1, \ldots, e_n)$; the value $\text{vol}(\text{conv}(e_1, \ldots, e_n)) = \frac{1}{n!}$ is computed by induction.

Theorem 40 (Urysohn's inequality)**.** *For every symmetric convex body $K \subset \mathbf{R}^n$, we have*

$$\text{vrad}(K) \leqslant w(K)$$

Proof. We use the following formula: if K is a convex body with 0 in the interior (which we can assume)

$$\text{vrad}(K) = \left(\int_{S^{n-1}} \|x\|_K^{-n}\, d\sigma(x)\right)^{1/n}. \tag{5.2}$$

To check (5.2) this formula, we integrate in polar coordinates

$$\begin{aligned}
\text{vol}(K) &= \int_{\mathbf{R}^n} \mathbf{1}_K(x)\,\mathrm{d}x \\
&= \lambda_n \int_{S^{n-1}} \int_0^\infty \mathbf{1}_K(r\theta) n r^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma(\theta) \\
&= \lambda_n \int_{S^{n-1}} \int_0^{\|\theta\|_K^{-1}} n r^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma(\theta) \\
&= \lambda_n \int_{S^{n-1}} \|\theta\|_K^{-n}\,\mathrm{d}\sigma(\theta)
\end{aligned}$$

for some constant $\lambda_n > 0$. The case $K = B_2^n$ shows that $\lambda_n = \text{vol}(B_2^n)$, proving (5.2). We now write, using Hölder inequality

$$\text{vrad}(K) = \left(\int_{S^{n-1}} \|x\|_K^{-n}\, d\sigma(x)\right)^{1/n} \geqslant \int_{S^{n-1}} \|x\|_K^{-1}\,\mathrm{d}\sigma(x) \geqslant \frac{1}{\int_{S^{n-1}} \|x\|_K\,\mathrm{d}\sigma(x)}.$$

If we now apply this inequality to K°, we get $1 \leqslant w(K)\,\text{vrad}(K^\circ)$. Combined with the Blaschke–Santalò inequality (which reads $\text{vrad}(K)\,\text{vrad}(K^\circ) \leqslant 1$), we obtain that $\text{vrad}(K) \leqslant w(K)$. □

Theorem 41. *Let $P \subset B_2^n$ be a polytope with N vertices. Then*

$$\text{vrad}(P) \leqslant C\sqrt{\frac{\log(N)}{n}}.$$

In particular, having $\mathrm{vrad}(P) \simeq 1$ requires an exponential number of vertices.

Proof. Let V be the set of vertices of P. Without loss of generality, we may assume that $V \subset S^{n-1}$ (check!). We then write

$$\mathrm{vrad}(P) \leqslant w(P) = \frac{1}{\kappa_n} \mathbf{E}\, w(P, G)$$

where G is a standard Gaussian vector in $\mathbf{R}^n$. The random variable $w(P, G)$ is the maximum of N random variables with distribution $N(0, 1)$, and the upper bound follows from Lemma 38 . □

The bound from Theorem 41 is not sharp when N is proportional to n. Here is an improvement in this range

Theorem 42. *Let $P \subset B_2^n$ be a polytope with λn vertices. Then*

$$\mathrm{vrad}(P) \leqslant C \frac{\lambda}{\sqrt{n}}.$$

Proof. We use Carathéodory's theorem (exo): any point $x \in P$ is a convex combination of at most $n + 1$ vertices. Geometrically, this means that P is the union of all simplices built on its vertices. Therefore, by the union bound,

$$\mathrm{vol}(P) \leqslant \binom{\lambda n}{n+1} v_n$$

where v_n is the maximal volume of a simplex inscribed inside B_2^n. Since $v_n \leqslant \frac{2 v_{n-1}}{n}$, we have $v_n \leqslant \frac{2^n}{n!}$, and therefore

$$\mathrm{vol}(P) \leqslant \binom{\lambda n}{n+1} \frac{2^n}{n!} \leqslant (\lambda n)^{n+1} \frac{(2e)^n}{n^n}.$$

We conclude by using the lower bound from (5.1). □

On can prove, under the hypotheses from Theorem 41, the upper bound $\mathrm{vol}(P) \leqslant C\sqrt{\frac{\log(N/n)}{n}}$, which is stronger than both Theorems 41 and 42.

5.2 Volume of the operator norm unit ball

As a preliminary to Gluskin's theorem, we need an estimate for the volume of the unit ball of the set of $n \times n$ matrices with respect to the operator norm. The operator norm on M_n is

$$\|A\|_{op} = \sup \left\{ \frac{|Ax|}{|x|} \ : \ x \neq 0 \right\}.$$

In order to do Euclidean geometry in M_n, we use the Hilbert–Schmidt inner product

$$\langle A, B\rangle = \mathrm{Tr}(AB^t)$$

and denote by $\|A\|_{HS} = \mathrm{Tr}(AA^t)^{1/2}$ the induced norm, called the Hilbert–Schmidt norm. We denote by B^n_{op} and B^n_{HS} the unit ball for the operator and Hilbert–Schmidt norms.

Proposition 43. *We have*

$$c\sqrt{n} \leqslant \left(\frac{\mathrm{vol}(B^n_{op})}{\mathrm{vol}(B^n_{HS})}\right)^{1/n^2} \leqslant \sqrt{n}.$$

Proof. The upper bound is easy and follows from the inequality $\|A\|_{op} \geqslant \frac{1}{\sqrt{n}}\|A\|_{HS}$, which can be seen by writing

$$\|A\|^2_{HS} = \left(\sum_{i=1}^n |Ae_i|^2\right)^{1/2} \leqslant \sqrt{n} \max_{1\leqslant i\leqslant n} |Ae_i| \leqslant \sqrt{n}\|A\|_{op}.$$

For the lower bound, we write ($S_{HS} \sim S^{n^2-1}$ being the Hilbert–Schmidt sphere, equipped with the uniform measure σ_{HS})

$$\frac{\mathrm{vol}(B^n_{op})}{\mathrm{vol}(B^n_{HS})} = \int_{S_{HS}} \|A\|_{op}^{-n^2} \,\mathrm{d}\sigma_{HS}(A) \geqslant \frac{1}{2} m^{-n^2},$$

where m is the median of $\|\cdot\|_{op}$ with respect to σ_{HS}. It remains to justify that $m \leqslant \frac{C}{\sqrt{n}}$. Since the function $\|\cdot\|_{op}$ is a 1-Lipschitz function on $(S_{HS}, \|\cdot\|_{HS})$, it follows by concentration of measure that its median and mean differ by at most $\frac{C}{\sqrt{n^2}} = \frac{C}{n}$. Therefore, we are reduced to show that

$$\int_{S_{HS}} \|A\|_{op} \,\mathrm{d}\sigma(A) \leqslant \frac{C}{\sqrt{n}}.$$

In view of (4.3), this is equivalent to the content of the next lemma. □

Lemma 44. *Let G a $n \times n$ matrix with independent $N(0,1)$ entries. Then*

$$\mathbf{E}\,\|G\|_{op} \leqslant C\sqrt{n}.$$

This is clearly sharp, since we have $\mathbf{E}\,\|G\|_{op} \geqslant \kappa_n$ by looking only at the first column.

Proof. Let $\mathcal{N}$ be a $\frac{1}{4}$-net in $(S^{n-1}, |\cdot|)$ with $\mathrm{card}\,\mathcal{N} \leqslant 9^n$. We have

$$\|G\|_{op} = \max_{x\in S^{n-1}} |Gx| = \max_{x,y\in S^{n-1}} \langle Gx, y\rangle.$$

Given x, $y \in S^{n-1}$, let x' and $y' \in \mathcal{N}$ such that $|x - x'| \leqslant \frac{1}{4}$ and $|y - y'| \leqslant \frac{1}{4}$. We have

$$\langle Gx, y \rangle \leqslant \langle Gx', y \rangle + |x - x'| \cdot |y| \cdot \|G\|_{op} \leqslant \langle Gx', y' \rangle + \frac{1}{2}\|G\|_{op}.$$

Taking the supremum over x, y gives

$$\|G\|_{op} \leqslant \max_{x', y' \in \mathcal{N}} \langle Gx', y' \rangle + \frac{1}{2}\|G\|_{op}$$

and thus

$$\mathbf{E}\,\|G\|_{op} \leqslant 2\,\mathbf{E} \max_{x', y' \in \mathcal{N}} \langle Gx', y' \rangle.$$

The right-hand side is the expectation of $N \leqslant 81^n$ random variables with distribution $N(0,1)$, and therefore

$$\mathbf{E}\,\|G\|_{op} \leqslant 2\sqrt{2\log(81^n)} = C\sqrt{n}.$$

□

5.3 Proof of Gluskin's theorem

Gluskin's theorem states that the diameter of the Banach–Mazur compactum BM_n is of order n.

Theorem 45. *There is a constant $c_0 > 0$ such that, for any dimension n, there exist symmetric convex bodies K_n, L_n in $\mathbf{R}^n$ such that $d_{BM}(K_n, L_n) \geqslant c_0 n$.*

Recall that

$$d_{BM}(K, L) = \inf \left\{ \frac{b}{a} \ : \ \exists T \in \mathsf{GL}_n(\mathbf{R}) \ \ aK \subset T(L) \subset bK \right\}$$

and that we can actually restrict to $T \in \mathsf{SL}_n^{\pm}(\mathbf{R})$, the set of $n \times n$ matrices with determinant equal to ± 1.

We will choose K_n and L_n at random. To motivate the proof, start with the following observation. It is trivial that $d_{BM}(B_1^n, B_1^n) = 1$. However, if $O \in \mathsf{O}(n)$ is chosen at random according to the Haar measure, We have $\mathbf{E}\,\|Oe_1\|_1 \sim c\sqrt{n}$ and therefore

$$\mathbf{P}\left(O(B_1^n) \subset c\sqrt{n} B_1^n\right)$$

tends to 0 as n grows. Therefore, the distance from B_1^n to itself, when computed at random, is of order n. Gluskin's idea is to exploit this phenomenon by considering random variants of B_1^n.

We consider

$$\mathcal{A}_n = \left\{ K \text{ of the form } \operatorname{conv}\{\pm x_i\}_{1 \leqslant i \leqslant 3n} \text{ with } x_i \in S^{n-1} \text{ and such that } K \supset \frac{1}{\sqrt{n}} B_2^n \right\}.$$

We define a $\mathcal{A}_n$-valued random variable by setting

$$K = \mathrm{conv}\{(\pm e_i)_{1 \leqslant i \leqslant n}, (\pm y_j)_{1 \leqslant j \leqslant 2n}\},$$

where $y_j \in S^{n-1}$ are i.i.d. with distribution σ. We say that K is a random convex body with distribution $\mathbf{P}_n$. We will prove Gluskin's theorem by showing that if K, L are independent random convex bodies with distribution $\mathbf{P}_n$, then for some c_0,

$$\mathbf{P}(d_{BM}(K, L) \geqslant c_0 n) \longrightarrow 1.$$

Proposition 46. *Fix $L \in \mathcal{A}_n$, and let K a random convex body with distribution $\mathbf{P}_n$. Then, for any $T \in \mathsf{SL}\pm_n(\mathbf{R})$ and $\rho \in (0, 1)$,*

$$\mathbf{P}\left(T(K) \subset \rho\sqrt{n}L\right) \leqslant \left(C_1\rho^2\right)^{n^2}$$

Proof. We generate K as $\mathrm{conv}\{\pm e_i, \pm y_j\}$ with (y_j) i.i.d. with distribution σ. If $T(K) \subset \rho\sqrt{n}L$, then $T(y_j) \in \rho\sqrt{n}L$ for every $j \in \{1, \dots, 2n\}$. These $2n$ events are independent, and therefore

$$\mathbf{P}(T(K) \subset \rho\sqrt{n}L) \leqslant \sigma\left(\left\{x \in S^{n-1} \ : \ T(x) \in \rho\sqrt{n}L\right\}\right)^{2n} \leqslant \sigma(S^{n-1} \cap \rho\sqrt{n}T^{-1}L)^{2n}.$$

Lemma 47. *If K_0 is a symmetric convex body in $\mathbf{R}^n$, then*

$$\sigma(S^{n-1} \cap K_0) \leqslant \frac{\mathrm{vol}(K_0)}{\mathrm{vol}(B_2^n)}.$$

Proof. Write

$$\sigma(S^{n-1} \cap K_0) = \frac{\{tx \ : \ t \in [0,1], \ x \in S^{n-1} \cap K_0\}}{\mathrm{vol}(B_2^n)} \leqslant \frac{\mathrm{vol}(K_0)}{\mathrm{vol}(B_2^n)}. \qquad \square$$

We continue the proof of Proposition 46. We have

$$\mathbf{P}(T(K) \subset \rho\sqrt{n}L) \leqslant \left(\frac{\mathrm{vol}(\rho\sqrt{n}L)}{\mathrm{vol}(B_2^n)}\right)^{2n} = \left(\rho\sqrt{n}\,\mathrm{vrad}(L)\right)^{2n^2}.$$

Since L is a polytope with $6n$ vertices, Theorem 42 implies that $\mathrm{vrad}(L) \leqslant C/\sqrt{n}$, and therefore

$$\mathbf{P}(T(K) \subset \rho\sqrt{n}L) \leqslant (C\rho)^{2n^2} = (C_1\rho^2)^{n^2}. \qquad \square$$

Proposition 46 shows that the event $T(K) \subset \rho\sqrt{n}L$ is unlikely for a fixed T. We are now going to use a net argument over $T \in \mathsf{SL}_n^{\pm}$.

Proposition 48. *Fix $L \in \mathcal{A}_n$, and denote*

$$\mathcal{M}_L = \{T \in \mathsf{M}_n(\mathbf{R}) \ : \ Te_i \in \sqrt{n}L \text{ for } 1 \leqslant i \leqslant n\}, \quad \text{(a convex set)}$$

$$\mathcal{T}_L = \mathcal{M}_L \cap \mathsf{SL}_n^{\pm}(\mathbf{R}).$$

For every $\varepsilon \in (0,1)$, $\mathcal{T}_L$ contains a ε-net (for $\|\cdot\|_{op}$) of cardinal at most $(C/\varepsilon)^{n^2}$.

Proof. Let $\mathcal{N} \subset \mathcal{T}_L$ a maximal ε-separated set for $\|\cdot\|_{op}$. Then $\mathcal{N}$ is a ε-net, and the balls $x_i + \frac{\varepsilon}{2}B_{op}^n$ for $x_i \in \mathcal{N}$ are disjoint and contained in $\mathcal{T}_L + \frac{\varepsilon}{2}B_{op}^n$. We claim that

$$\mathcal{T}_L + \frac{\varepsilon}{2}B_{op}^n \subset \left(1 + \frac{\varepsilon}{2}\right)\mathcal{M}_L.$$

Indeed, we have $\mathcal{T}_L \subset \mathcal{M}_L$ (obvious) and if $T \in B_{op}^n$, then $Te_i \in B_2^n \subset \sqrt{n}L$, so $B_{op}^n \subset \mathcal{M}_L$.

Comparing volumes gives

$$\mathrm{card}(\mathcal{N})\,\mathrm{vol}\left(\frac{\varepsilon}{2}B_{op}^n\right) \leqslant \left(1+\frac{\varepsilon}{2}\right)^{n^2}\mathrm{vol}(\mathcal{M}_L).$$

By Fubini's theorem, we have $\mathrm{vol}_{n^2}(\mathcal{M}_L) = \mathrm{vol}_n(\sqrt{n}L)^n$. As we already observed, $\mathrm{vrad}(L) \leqslant C/\sqrt{n}$, so $\mathrm{vol}(L) \leqslant (C/n)^n$ and $\mathrm{vol}(\mathcal{M}_L) \leqslant (C/\sqrt{n})^{n^2}$. On the other hand, we know from Proposition 43 that $\mathrm{vol}(B_{op}^n) \geqslant (C/\sqrt{n})^{n^2}$. This gives

$$\mathrm{card}(\mathcal{N}) \leqslant \left(\frac{3}{\varepsilon}\right)^{n^2}\frac{\mathrm{vol}(\mathcal{M}_L)}{\mathrm{vol}(B_{op}^n)} \leqslant \left(\frac{C}{\varepsilon}\right)^{n^2}. \qquad \square$$

We have now all the ingredients needed to prove Gluskin's theorem. Fix $L \in \mathcal{A}_n$, and let K be a random convex body.

Lemma 49. *If $\varepsilon < \rho < 1$, then*

$$\mathbf{P}\left(\exists T \in \mathsf{SL}_n^{\pm}(\mathbf{R}) \ : \ T(K) \subset (\rho-\varepsilon)\sqrt{n}L\right) \leqslant \left(\frac{C\rho^2}{\varepsilon}\right)^{n^2}$$

Proof. Let $\mathcal{N}$ be a ε-net in $\mathcal{T}_L$ given by Proposition 48, with $\mathrm{card}\,\mathcal{N} \leqslant (C/\varepsilon)^{n^2}$. Assume that there exists $T \in \mathsf{SL}_n^{\pm}(\mathbf{R})$ such that $T(K) \subset (\rho-\varepsilon)\sqrt{n}L$. Then $T(e_i) \in \sqrt{n}L$ for every i, and therefore $T \in \mathcal{T}_L$. Choose $T' \in \mathcal{N}$ such that $\|T - T'\|_{op} \leqslant \varepsilon$. For every $x \in K$, we have

$$\|T'x\|_L \leqslant \|Tx\|_L + \|(T-T')x\|_L \leqslant (\rho-\varepsilon)\sqrt{n} + \sqrt{n}|(T-T')x| \leqslant \rho\sqrt{n}.$$

The problem has been discretized: we have

$$\begin{aligned}
\mathbf{P}\left(\exists T \in \mathsf{SL}_n^{\pm}(\mathbf{R}) \ : \ T(K) \subset (\rho-\varepsilon)\sqrt{n}L\right) &\leqslant \mathbf{P}\left(\exists T' \in \mathcal{N}_L \ : \ T'(K) \subset \rho\sqrt{n}L\right) \\
&\leqslant \operatorname{card}(\mathcal{N}_L) \sup_{T' \in \mathcal{N}_L} \mathbf{P}\left(T'(K) \subset \rho\sqrt{n}L\right) \\
&\leqslant \left(\frac{C}{\varepsilon}\right)^{n^2} \left(C_1\rho^2\right)^{n^2} \\
&\leqslant \left(\frac{C\rho^2}{\varepsilon}\right)^{n^2}
\end{aligned}$$

as needed. □

We now choose $\rho = \frac{1}{4C}$ and $\varepsilon = \frac{\rho}{2}\frac{1}{8C}$, so that $\frac{C\rho^2}{\varepsilon} = \frac{1}{2}$. We have shown that for a fixed $L \in \mathcal{A}_n$ and a random K,

$$\mathbf{P}\left(\exists T \in \mathsf{SL}_n^{\pm}(\mathbf{R}) \ : \ T(K) \subset \frac{1}{8C}\sqrt{n}L\right) \leqslant 2^{-n^2}.$$

Let now K and L be independent random convex bodies. By conditioning,

$$\mathbf{P}\left(\exists T \in \mathsf{SL}_n^{\pm}(\mathbf{R}) \ : \ K \subset \frac{\sqrt{n}}{8C}T(L)\right) \leqslant 2^{-n^2}.$$

$$\mathbf{P}\left(\exists T \in \mathsf{SL}_n^{\pm}(\mathbf{R}) \ : \ T(L) \subset \frac{\sqrt{n}}{8C}K\right) \leqslant 2^{-n^2}.$$

With probability at least $1 - 2 \cdot 2^{-n^2}$, if $T \in \mathsf{SL}_n^{\pm}(\mathbf{R})$ and $a, b > 0$ satisfy $aK \subset T(L) \subset bK$, then $b > \sqrt{n}/8C$ and $a^{-1} > \sqrt{n}/8C$, so $b/a \geqslant n/(64C^2)$. This shows that

$$\mathbf{P}\left(d_{BM}(K,L) \leqslant \frac{n}{64C^2}\right) \leqslant 1 - 2^{1-n^2},$$

proving Theorem 45.

Chapter 6

Gaussian processes

By a *stochastic process* we just mean a collection $(X_t)_{t\in T}$ of random variables. We say that $(X_t)_{t\in T}$ is a (centered) *Gaussian process* if every linear combination

$$\sum_{t\in T} \lambda_t X_t$$

has a centered Gaussian distribution $N(0,\sigma^2)$ for some $\sigma \geqslant 0$.

When $(X_t)_{t\in T}$ is a Gaussian process, the index set T can be equipped with a distance induced by the L^2 norm: for $s,t \in T$

$$d(s,t) = \left(\mathbf{E}\left[|X_s - X_t|^2\right]\right)^{1/2}.$$

Example of Gaussian process can be constructed as follows: consider any subset $T \subset \mathbf{R}^n$, and set

$$X_t = \langle G, t\rangle$$

where G is a $N(0,\mathrm{Id})$ Gaussian random vector. This example describes the general case, at least when T is finite. Indeed, given a Gaussian process $(X_t)_{t\in T}$ with T finite, we may identify the subspace $\mathrm{span}\{X_t \ : \ t\in T\} \subset L^2(\Omega)$ with the Euclidean space $(\mathbf{R}^n, |\cdot|)$ for some n. This induces a map $\phi : T \to \mathbf{R}^n$. If we set $Y_t := \langle G, \phi(T)\rangle$ with G as above, we check that

$$\mathbf{E}\,Y_t^2 = |\phi(t)|^2 = \mathbf{E}\,X_t^2,$$

$$2\,\mathbf{E}\,Y_sY_t = \mathbf{E}\,Y_s^2 + \mathbf{E}\,Y_t^2 - \mathbf{E}(Y_s - Y_t)^2 = \mathbf{E}\,X_s^2 + \mathbf{E}\,X_t^2 - \mathbf{E}(X_s - X_t)^2 = 2\,\mathbf{E}\,X_sX_t.$$

Since the distribution of a centered Gaussian process is characterized by the covariance matrix, the vectors $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ have the same distribution.

The goal of this chapter is to give estimates on the quantity

$$\mathbf{E} \sup_{t\in T} X_t$$

in terms of the geometry of the metric space (T,d). In full generality, measurability issues could arise, but in practice we will always reduce to the case when T is finite.

6.1 Comparison inequalities

Lemma 50 (Slepian's lemma)**.** *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be Gaussian processes, with T finite. Assume that $\mathbf{E}\, X_t^2 = \mathbf{E}\, Y_t^2$ for every $t \in T$, and also that for every s, t*

$$\|X_s - X_t\|_{L^2} \leqslant \|Y_s - Y_t\|_{L^2}.$$

Then, for any real numbers (λ_t),

$$\mathbf{P}\left(\exists t \ : \ X_t \geqslant \lambda_t\right) \leqslant \mathbf{P}\left(\exists t \ : \ Y_t \geqslant \lambda_t\right), \tag{6.1}$$

which implies in particular that

$$\mathbf{E}\max_{t\in T} X_t \leqslant \mathbf{E}\max_{t\in T} Y_t. \tag{6.2}$$

We first explain the last part of the lemma. It is useful to know about stochastic domination. Given random variables X, Y, the following are equivalent (check!) and we say that Y dominates X

1. For every $\lambda \in \mathbf{R}$, $\mathbf{P}(X \geqslant \lambda) \leqslant \mathbf{P}(Y \geqslant \lambda)$,
2. For every measurable non-decreasing function $f : \mathbf{R} \to \mathbf{R}$ such that $f(X)$ and $f(Y)$ are integrable, we have $\mathbf{E}[f(X)] \leqslant \mathbf{E}[f(Y)]$.
3. There are random variables X' and Y' defined on a common probability space, such that X and X' have the same law, Y and Y' have the same law, and $\mathbf{P}(X' \leqslant Y') = 1$.

It is then easy to check that (6.1) implies that $\max Y_t$ dominates $\max X_t$, and (6.2) follows (check!).

We now state a generalization of Slepian's lemma. It is more complicated to state, but not harder to prove. Slepian's lemma appears at the special case where each set T_s is a singleton.

Lemma 51 (Gordon's lemma)**.** *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be Gaussian processes, with T finite. Assume that T is written as a partition $T = \bigcup_{s\in S} T_s$, and for $t \in T$ denote by $s(t)$ the unique s such that $t \in T_s$. We assume that $\mathbf{E}\, X_t^2 = \mathbf{E}\, Y_t^2$ for every $t \in T$, and that for t, $t' \in T$*

$$\begin{aligned} \|X_t - X_{t'}\|_{L^2} &\leqslant \|Y_t - Y_{t'}\|_{L^2} \quad \text{if } s(t) \neq s(t') \\ \|X_t - X_{t'}\|_{L^2} &\geqslant \|Y_t - Y_{t'}\|_{L^2} \quad \text{if } s(t) = s(t') \end{aligned}$$

Then, for every real numbers (λ_t),

$$\mathbf{P}\left(\bigcup_{s\in S}\bigcap_{t\in T_s}\{X_t \geqslant \lambda_t\}\right) \leqslant \mathbf{P}\left(\bigcup_{s\in S}\bigcap_{t\in T_s}\{Y_t \geqslant \lambda_t\}\right), \tag{6.3}$$

which implies in particular that

$$\mathbf{E}\max_{s\in S}\min_{t\in T_s} X_t \leqslant \mathbf{E}\max_{s\in S}\min_{t\in T_s} Y_t$$

It is useful to remark that the process $(-X_t)$, $(-Y_t)$ also satisfy the hypothesis of Gordon's lemma, and therefore it also holds that

$$\mathbf{E}\min_{s\in S}\max_{t\in T_s} X_t \geqslant \mathbf{E}\min_{s\in S}\max_{t\in T_s} Y_t.$$

Proof. We note that, taking complements, (6.3) is equivalent to

$$\mathbf{E}\left[\prod_{s\in S}\left(1-\prod_{t\in T_s}\mathbf{1}_{\{X_t\geqslant\lambda_t\}}\right)\right] \geqslant \mathbf{E}\left[\prod_{s\in S}\left(1-\prod_{t\in T_s}\mathbf{1}_{\{Y_t\geqslant\lambda_t\}}\right)\right].$$

We show a functional version of this inequality: whenever (f_t) are non-decreasing functions with values in $[0,1]$,

$$\mathbf{E}\left[\prod_{s\in S}\left(1-\prod_{t\in T_s} f_t(X_t)\right)\right] \geqslant \mathbf{E}\left[\prod_{s\in S}\left(1-\prod_{t\in T_s} f_t(Y_t)\right)\right],$$

the previous inequality corresponding to $f_t = \mathbf{1}_{[\lambda_t,+\infty)}$. We can now assume that each function f_t is of class C^2. If we introduce the function $F:\mathbf{R}^T\to\mathbf{R}$ defined by

$$F((x_t)_{t\in T}) = \prod_{s\in S}\left(1-\prod_{t\in T_s} f_t(x_t)\right),$$

we are reduced to showing that $\mathbf{E}\,F(X_t)\geqslant\mathbf{E}\,F(Y_t)$. We observe the following: for $u,v\in T$,

$$\begin{cases}\partial^2_{uv}F\geqslant 0 & \text{if } s(u)\neq s(v)\\ \partial^2_{uv}F\leqslant 0 & \text{if } s(u)=s(v) \text{ and } u\neq v.\end{cases}$$

We interpolate between (X_t) and (Y_t) as follows. First, we may assume that (X_t) and (Y_t) are independent (check!). Next, define for $\theta\in[0,\pi/2]$,

$$W_t(\theta) = \cos(\theta)X_t + \sin(\theta)Y_t$$

so that $W_t(0)=X_t$ and $W_t(\pi/2)=Y_t$. If we consider the function $\Phi(\theta)=\mathbf{E}[F(W_t(\theta))]$, it is enough to show that $\Phi'\leqslant 0$ on $[0,\pi/2]$. For a fixed $\theta\in[0,\pi/2]$, we compute

$$\Phi'(\theta) = \mathbf{E}\sum_{u\in T}\partial_u F(W_t(\theta))W_u'(\theta),$$

where $W_t'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} W_t(\theta) = -\sin(\theta) X_t + \cos(\theta) Y_t$. We now also fix $u \in T$. We use the following formula (check!): if (G, H) is a pair of jointly Gaussian variables, we may write $G = \alpha H + Z$ for $\alpha \in \mathbf{R}$ and Z a random variable independent from Z (and we then have $\alpha = \frac{\mathbf{E}[GH]}{\mathbf{E}[H^2]}$).

Therefore, for every $t \in T$, we may write

$$W_t(\theta) = \alpha_t W_u'(\theta) + Z_t$$

with Z_t independent from $W_t'(\theta)$. The real number α_t has the same sign as

$$\mathbf{E}[W_t(\theta) W_t'\theta)] = \cos(\theta)\sin(\theta)\left(\mathbf{E}[Y_t Y_u] - \mathbf{E}[X_t X_u]\right).$$

From our hypothesis, we see that $\alpha_t \geqslant 0$ if $s(t) = s(u)$ and $\alpha_t \leqslant 0$ if $s(t) \neq s(u)$. Moreover, $\alpha_u = 0$.

We write

$$\Phi'(\theta) = \sum_{u \in T} \mathbf{E}_{\omega \in \Omega} \underbrace{W_u'(\theta)(\omega) \partial_u F((\alpha_t W_u'(\theta)(\omega) + Z_t(\omega))}_{h_{u,\omega}((\alpha_t)_{t \in T})}.$$

We now focus on the quantity $h_{u,\omega}$ from the previous equation, which we think of as a function of the variables $(\alpha_t)_{t \in T}$. We have

$$\partial_t h_{u,\omega} = (W_u')^2 \partial^2_{ut}(\alpha_t W_u' + Z_t) \begin{cases} \geqslant 0 & \text{if } s(t) \neq s(u), \\ \leqslant 0 & \text{if } s(t) = s(u), t \neq u. \end{cases}$$

Since α_t has a sign opposed to $\partial_t h_{u,\omega}$, it follows that $h_{u,\omega}((\alpha_t)_{t \in T}) \leqslant h_{u,\omega}(0, \dots, 0)$. Therefore, we have

$$\Phi'(\theta) \leqslant \sum_{u \in T} \mathbf{E}\left[W' u(\theta) \partial_u F(Z_t)\right] = 0,$$

where the last equality follows from the independence of Z_t and W_u'. The proof is therefore complete. □

Here is a variant on Slepian's lemma.

Lemma 52 (Fernique's lemma)**.** *Let $(X_t)_{t \in T}$, $(Y_t)_{t \in T}$ be Gaussian processes, with T finite. Assume that $\|X_s - X_t\|_{L^2} \leqslant \|Y_s - Y_t\|_{L^2}$ for every s, $t \in T$. Then*

$$\mathbf{E} \max_{t \in T} X_t \leqslant \mathbf{E} \max_{t \in T} Y_t.$$

It is clear that stochastic domination does not hold without the hypothesis $\mathbf{E}\, X_t^2 = \mathbf{E}\, Y_t^2$ (consider the case of T being a singleton).

Proof. Let $Z \sim N(0,1)$ be a random variable independent from (X_t) and (Y_t). For $\varepsilon \in (0,1)$ and $R > 0$ large enough, we define

$$\overline{X}_t = (1-\varepsilon)X_t + \alpha_t Z,$$

$$\overline{Y}_t = Y_t + \beta_t Z,$$

where α_t and β_t are chosen so that $\mathbf{E}\,\overline{X}_t^2 = \mathbf{E}\,\overline{Y}_t^2 = R^2$. In formulas, we have (as $R \to \infty$)

$$\alpha_t = \sqrt{R^2 - (1-\varepsilon)^2\,\mathbf{E}\,X_t^2} = R - \frac{(1-\varepsilon)^2\,\mathbf{E}\,X_t^2}{2R} + o(1/R),$$

$$\beta_t = \sqrt{R^2 - \mathbf{E}\,Y_t^2} = R - \frac{\mathbf{E}\,Y_t^2}{2R} + o(1/R).$$

We have

$$\|\overline{X}_s - \overline{X}_t\|_{L^2}^2 = (1-\varepsilon)^2\|X_s - X_t\|_{L^2}^2 + (\alpha_s - \alpha_t)^2 \underset{R\to\infty}{\to} (1-\varepsilon)^2\|X_s - X_t\|_{L^2}^2,$$

$$\|\overline{Y}_s - \overline{Y}_t\|_{L^2}^2 = \|Y_s - Y_t\|_{L^2}^2 + (\beta_s - \beta_t)^2 \underset{R\to\infty}{\to} \|Y_s - Y_t\|_{L^2}^2.$$

In particular, for R large enough, we have $\|\overline{X}_s - \overline{X}_t\|_{L^2} \leqslant \|\overline{Y}_s - \overline{Y}_t\|_{L^2}$ for every s,t. We may therefore apply Slepian's lemma to the processes $(\overline{X}_t)$ and $(\overline{Y}_t)$ and conclude that $\mathbf{E}\max \overline{X}_t \leqslant \mathbf{E}\max \overline{Y}_t$. Note that

$$\mathbf{E}\max_{t\in T} \overline{X}_t = \mathbf{E}\max_{t\in T}(\overline{X}_t - RZ) = (1-\varepsilon)\,\mathbf{E}\max_{t\in T} X_t + O(1/R),$$

$$\mathbf{E}\max_{t\in T} \overline{Y}_t = \mathbf{E}\max_{t\in T}(\overline{Y}_t - RZ) = \mathbf{E}\max_{t\in T} Y_t + O(1/R),.$$

Letting $R \to \infty$ gives $(1-\varepsilon)\,\mathbf{E}\max X_t \leqslant \mathbf{E}\max Y_t$, and the result follows by taking ε to zero. □

Nice applications of Slepian's lemma arise when considering random matrices. Here is an example, which improves on 44. The constant 2 can be shown to be sharp.

Proposition 53. *Let G be a $n \times n$ matrix with independent $N(0,1)$ entries. Then $\mathbf{E}\,\|G\|_{op} \leqslant 2\sqrt{n}$.*

Proof. We consider two Gaussian processes indexed by $S^{n-1} \times S^{n-1}$

$$X_{(x,y)} = \langle Gx, y\rangle,$$

$$Y_{(x,y)} = \langle g_1, x\rangle + \langle g_2, x\rangle,$$

with g_1 and g_2 independent $N(0, \mathrm{Id}_n)$ Gaussian vectors. We note that

$$\mathbf{E}\,\|G\|_{op} = \mathbf{E}\max_{(x,y)\in S^{n-1}\times S^{n-1}} X_{(x,y)}.$$

We claim that for every x, y, x', y' in S^{n-1},

$$\|X_{(x,y)} - X_{(x',y')}\|_{L^2} \leqslant \|Y_{(x,y)} - Y_{(x',y')}\|_{L^2}. \tag{6.4}$$

For every finite subset $T \subset S^{n-1} \times S^{n-1}$, we apply Slepian's lemma to the processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$. When T ranges over all finite subsets of $S^{n-1} \times S^{n-1}$, this gives (check!)

$$\mathbf{E}\,\|G\|_{op} = \mathbf{E} \max_{(x,y)\in S^{n-1}\times S^{n-1}} X_{(x,y)} \leqslant \mathbf{E} \max_{(x,y)\in S^{n-1}\times S^{n-1}} Y_{(x,y)} = \mathbf{E}\,[|g_1| + |g_2|] = 2\kappa_n \leqslant \sqrt{n}.$$

It remains to justify (6.4). We compute that

$$\|X_{(x,y)} - X_{(x',y')}\|_{L^2}^2 = \sum_{i,j}(x_i y_j - x_i' y_j')^2 = 2 - \sum_{i,j} x_i y_j x_i' y_j' = 2 - \langle x, x'\rangle\langle y, y'\rangle$$

$$\|Y_{(x,y)} - Y_{(x',y')}\|_{L^2}^2 = \sum_i (x_i - x_i')^2 + \sum_j (y_j - y_j')^2 = 2 - \langle x, x'\rangle + 2 - \langle y, y'\rangle.$$

Since $2(1 - \langle x, x'\rangle)(1 - \langle y, y'\rangle) \geqslant 0$, the inequality follows. □

A similar argument applies to rectangular matrices. In that case, extra information can be obtain by using Gordon's lemma.

Proposition 54. *Let G be a $m \times n$ matrix with independent $N(0,1)$ entries, for $n \leqslant m$. Consider G as a linear map from $\mathbf{R}^n$ to $\mathbf{R}^m$. Then,*

$$\sqrt{m} - \sqrt{n} \leqslant \mathbf{E} \min_{x\in S^{n-1}} |Gx| \leqslant \mathbf{E} \max_{x\in S^{n-1}} |Gx| \leqslant \sqrt{n} + \sqrt{m}.$$

Proof. We consider the Gaussian processes indexed by $S^{n-1} \times S^{m-1}$

$$X_{(x,y)} = \langle Gx, y\rangle,$$

$$Y_{(x,y)} = \langle g, x\rangle + \langle h, y\rangle,$$

with $g \sim N(0, \mathrm{Id}_n)$ and $h \sim (O, \mathrm{Id}_m)$. We have, as in the previous proof,

$$\|Y_{(x,y)} - Y_{(x',y')}\|_{L^2}^2 - \|X_{(x,y)} - X_{(x',y')}\|_{L^2}^2 = 2(1 - \langle x, x'\rangle)(1 - \langle y, y'\rangle).$$

It follows that the hypotheses of Gordon's lemma are satisfied if we equip the index set with the partition

$$S^{n-1} \times S^{m-1} = \bigcup_{s\in S^{n-1}} \{s\} \times S^{m-1}.$$

Gordon's lemma implies that (check!)

$$\kappa_m - \kappa_n \leqslant \mathbf{E} \min_{x\in S^{n-1}} |Gx| \leqslant \mathbf{E} \max_{x\in S^{n-1}} |Gx| \leqslant \kappa_m + \kappa_n$$

and (not so easy) considerations from calculus show that $\kappa_m - \kappa_n \geqslant \sqrt{m} - \sqrt{n}$ whenever $m \geqslant n$. □

6.2 Sudakov inequalities

Let $(X_t)_{t\in T}$ be a Gaussian process. For $\varepsilon > 0$, denote by $N(T, d, \varepsilon)$ the covering number of the metric space (T, d).

Proposition 55 (Sudakov inequality)**.** *Let $(X_t)_{t\in T}$ be a centered Gaussian process. Then, for every $\varepsilon > 0$,*

$$\mathbf{E}\sup_{t\in T} X_t \geqslant c\varepsilon\sqrt{\log N(T, d, \varepsilon)}.$$

Proof. Let $N = N(T, d, \varepsilon)$. By (3.3), there is a subset $(t_i)_{1\leqslant i\leqslant N}$ of T such that $d(t_i, t_j) \geqslant \varepsilon$ whenever $i \neq j$. Let $(Z_i)_{1\leqslant i\leqslant N}$ be i.i.d. $N(0, 1)$ random variables, and $Y_i = \frac{\varepsilon}{\sqrt{2}} Z_i$. For $i \neq j$, we have $\|Z_i - Z_j\|_{L^2} = \sqrt{2}$ and therefore $\|Y_i - Y_j\|_{L^2} = \varepsilon \leqslant \|X_{t_i} - X_{t_j}\|_{L^2}$. By Fernique's lemma, we have

$$\mathbf{E}\sup_{1\leqslant i\leqslant N} Y_i \leqslant \mathbf{E}\sup_{1\leqslant i\leqslant N} X_{t_i} \leqslant \mathbf{E}\sup_{t\in T} X_t.$$

We know from Lemma 38 that the left-hand-side is greater that $c\varepsilon\sqrt{\log N}$. □

As a corollary, we obtain upper bounds on the covering number of convex bodies. Given convex bodies $K, L \subset \mathbf{R}^n$, denote by $N(K, L, \varepsilon)$ the minimal number of translates of εL needed to cover K. In other words,

$$N(K, L, \varepsilon) = \inf\left\{ N \ : \ \exists x_1, \ldots, x_N \in K \ : \ K \subset \bigcup_{i=1}^{N} x_i + \varepsilon L \right\}.$$

Corollary 56. *Let $K \subset \mathbf{R}^n$ be a convex body. Then*

$$\log N(K, B_2^n, \varepsilon) \leqslant C\frac{nw(K)^2}{\varepsilon^2}$$

Proof. Apply Sudakov's inequality to the Gaussian process $(X_t)_{t\in T}$ defined by $X_t = \langle G, t\rangle$, where $T = K$ and G is a standard Gaussian vector in $\mathbf{R}^n$. Note that the metric space (T, d) can be identified with $(K, |\cdot|)$, and that

$$\mathbf{E}\sup_{t\in T} X_t = \kappa_n w(K).$$

□

It is conjectured that the covering numbers of convex bodies satisfy the following (approximate) duality property: if K, L are symmetric convex bodies in $\mathbf{R}^n$, then do we have

$$\log N(L^\circ, K^\circ, C\varepsilon) \leqslant C \log N(K, L, \varepsilon) \ ? \tag{6.5}$$

The inequality (6.5) (which is known to be true when $L = B_2^n$, but this is not an easy result) implies a dual version of Sudakov's inequality.

Proposition 57 (Dual Sudakov inequality)**.** *If $K \subset \mathbf{R}^n$ is a symmetric convex body, then*

$$\log N(B_2^n, K^\circ, \varepsilon) \leqslant C \frac{n w(K)^2}{\varepsilon^2}.$$

Proof. Let

$$w_g(K) = \kappa_n w(K) = \int_{\mathbf{R}^n} \sup_{x \in K} \langle x, y \rangle \, \mathrm{d}\gamma_n(y)$$

be the Gaussian mead width of K. We may assume that $w_g(K) = 1$ (otherwise, replace K by λK for $\lambda = w_g(K)^{-1}$). We have to show that $\log N(rB_2^n, K^\circ) \leqslant Cr^2$ for $r > 0$.

Let $x_1, \dots, x_N \in rB_2^n$ such that the sets $(x_i + 2K^\circ)$ are disjoint. We can remark that since $w_g(K) = 1$, we have $\gamma_n(2K^\circ) \geqslant \frac{1}{2}$ by Markov's inequality. Moreover, using symmetry of K°, we have

$$\begin{aligned} \gamma_n(x_i + 2K^\circ) &= \frac{\gamma_n(x_i + 2K^\circ) + \gamma_n(-x_i + 2K^\circ)}{2} \\ &= \int_{2K^\circ} \frac{\Phi(x + x_i) + \Phi(x - x_i)}{2} \, \mathrm{d}x \end{aligned}$$

where $\Phi(x) = \frac{1}{(2\pi)^{n/2}} \exp(-|x|^2/2)$ is the Gaussian density. We have, using convexity of the exponential function,

$$\begin{aligned} \frac{\Phi(x + x_i) + \Phi(x - x_i)}{2} &\geqslant \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{|x + x_i|^2}{4} - \frac{|x - x_i|^2}{4} \right) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{|x|^2}{2} - \frac{|x_i|^2}{2} \right) \\ &\geqslant \Phi(x) \exp(-r^2/2). \end{aligned}$$

Integrating over $2K^\circ$ gives

$$\gamma_n(x_i + 2K^\circ) \geqslant e^{-r^2/2} \gamma_n(2K^\circ) \geqslant \frac{1}{2} e^{-r^2/2}.$$

Since the sets $(x_i + 2K^\circ)$ are disjoint, it follows that $\frac{1}{2} e^{-r^2/2} N \leqslant 1$, completing the proof. □

6.3 Dudley inequality

Dudley's inequality is an upper bound on the expected supremum of a Gaussian process, in terms of covering numbers. It actually holds true, with the same proof, for the larger class of *subGaussian* processes which we now introduce.

A centered stochastic process (X_t) indexed by a metric space (T, d) is subGaussian with constant $\alpha > 0$ if for every $s, t \in T$ and $x > 0$,

$$\mathbf{P}(X_s - X_t > x) \leqslant 2 \exp\left(-\alpha \frac{x^2}{d(s,t)^2} \right).$$

If $(X_t)_{t\in T}$ is a Gaussian process and d is the distance on T induced from the L^2 norm, then (X_t) is subGaussian with constant $\frac{1}{2}$: if s, $t \in T$, then $\frac{X_s - X_t}{d(s,t)} \sim N(0,1)$ and we use the fact that a $N(0,1)$ random variable X satisfies

$$\mathbf{P}(X > x) \leqslant \frac{1}{2}\exp(-x^2/2).$$

Theorem 58 (Dudley's inequality)**.** *Let (X_t) be a centered subGaussian process with constant α. Then*

$$\mathbf{E}\sup_{t\in T} X_t \leqslant \frac{C}{\sqrt{\alpha}} \int_0^\infty \sqrt{\log N(T,d,\varepsilon)}\,\mathrm{d}\varepsilon. \tag{6.6}$$

If the metric space T is bounded (which is always the case in applications), then $N(T,d,\varepsilon) = 1$ for ε larger that ε_0 enough and therefore the integral can be taken on $[0,\varepsilon_0]$.

Proof. We actually show the equivalent bound

$$\mathbf{E}\sup_{t\in T} X_t \leqslant \frac{C}{\sqrt{\alpha}} \sum_{k\in\mathbf{Z}} 2^{-k}\sqrt{\log N(T,d,2^{-k})}. \tag{6.7}$$

If I denotes the integral in (6.6) and S denotes the series in (6.7), then $S \leqslant I \leqslant 2S$. Indeed, write

$$I = \sum_{k\in\mathbf{Z}} \int_{2^k}^{2^{k+1}} \sqrt{\log N(T,d,\varepsilon)}\,\mathrm{d}\varepsilon$$

and use the fact that the function $\varepsilon \mapsto N(T,d,\varepsilon)$ is nonincreasing.

When proving Dudley's theorem, we can assume that T is finite (check!) and that $\alpha = 1$ (by homogeneity: if (X_t) is subGaussian with constant α, then (cX_t) is subGaussian with constant $\alpha/\sqrt{c}$).

For $k \in \mathbf{Z}$, set $\varepsilon_k = 2^{-k}$, and let $\mathcal{N}_k$ be a ε_k-net in (T,d) such that $\operatorname{card}(\mathcal{N}_k) = N(T,d,\varepsilon_k)$. We also write $k_{\max}$ for the minimal k such that $\mathcal{N}_k = T$ and $k_{\min}$ for the maximal k such that $\operatorname{card}(\mathcal{N}_k) = 1$. Therefore $\mathcal{N}_{k_{\min}} = \{t_0\}$.

For $t \in T$ and $k \in \mathbf{Z}$, let $\pi_k(t) \in \mathcal{N}_k$ such that $d(t,\pi_k(t)) \leqslant \varepsilon_k$. We have

$$\mathbf{E}\sup_{t\in T} X_t = \mathbf{E}\sup_{t\in T}(X_t - X_{t_0}).$$

The idea will be to use *chaining*: write

$$X_t - X_{t_0} = \sum_{k=k_{\min}}^{k_{\max}-1} X_{\pi_{k+1}(t)} - X_{\pi_k(t)}$$

and therefore

$$\mathbf{E}\sup_{t\in T}(X_t - X_{t_0}) \leqslant \sum_{k=k_{\min}}^{k_{\max}-1} \mathbf{E}\sup_{t\in T}(X_{\pi_{k+1}(t)} - X_{\pi_k(t)}).$$

We now focus on the quantity $\mathbf{E}\sup_{t\in T}(X_{\pi_{k+1}(t)}-X_{\pi_k(t)})$, for fixed k. This is the supremum of at most $\mathrm{card}(\mathcal{N}_k)\,\mathrm{card}(\mathcal{N}_{k+1})$ random variables, each satisfying the subGaussian estimate

$$\mathbf{P}(X_{\pi_{k+1}(t)}-X_{\pi_k(t)}) > x) \leqslant 2\exp\left(-\frac{x^2}{d(\pi_{k+1}(t),\pi_k(t))^2}\right) \leqslant 2\exp\left(-\frac{x^2}{(2\varepsilon_k)^2}\right).$$

We have the following lemma (check!)

Lemma 59. *Let $Y_1,\dots,Y_N$ be random variables satisfying $\mathbf{P}(Y_i > x) \leqslant 2\exp(-x^2/\beta^2)$ for $N \geqslant 2$. Then $\mathbf{E}\max(Y_1,\dots,Y_N) \leqslant C\beta\sqrt{\log N}$.*

It follows that

$$\mathbf{E}\sup_{t\in T}(X_{\pi_{k+1}(t)}-X_{\pi_k(t)}) \leqslant C\varepsilon_k\sqrt{\log\left(\mathrm{card}(\mathcal{N}_k)\,\mathrm{card}(\mathcal{N}_{k+1})\right)} \leqslant C\varepsilon_k\sqrt{\log N(T,d,\varepsilon_{k+1}}.$$

Combining all the estimates gives

$$\mathbf{E}\sup_{t\in T} X_t \leqslant C\sum_{k=k_{\min}}^{k_{\max}} 2^{-k-1}\sqrt{\log N(T,d,2^{-k-1})}$$

and (6.7) follows. □

As an application of Dudley's inequality, we prove a uniform law of large numbers. Consider an integrable function $f:[0,1]\to\mathbf{R}$. If (Z_i) are i.i.d. random variables uniformly distributed on $[0,1]$, the by the law of large numbers

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f(Z_i) = \int_0^1 f(x)\,\mathrm{d}x.$$

Moreover, the error is of order $O(1/\sqrt{n})$ when $f\in L^2$. Can we hope for the error to be small simultaneously for every f? This is clear not possible: given samples $(Z_1,\dots,Z_n)$, one may engineer a function f for which the empirical mean is arbitrary large from the limit. However, this becomes true if we impose some mild regularity on f, for example being Lipschitz.

Theorem 60. *Let $\mathcal{F}$ be the family of L-Lipschitz function from $[0,1]$ to $\mathbf{R}$. Then, if (Z_i) are i.i.d. uniformly distributed on $[0,1]$,*

$$\mathbf{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f(Z_i) - \int_0^1 f(x)\,\mathrm{d}x\right| \leqslant \frac{CL}{\sqrt{n}}.$$

Proof. We may assume $L = 1$ by homogeneity. We may also consider equivalently the subclass $\mathcal{F}_0$ of functions with integral equal to 0. Consider the process $(X_f)_{f\in\mathcal{F}_0}$ defined by

$$X_f = \frac{1}{n}\sum_{i=1}^{n} f(Z_i).$$

We recall the classical Hoeffding inequality

Lemma 61 (Hoeffding's inequality)**.** *Let $Y_1, \ldots, Y_n$ be independent random variables, such that Y_i takes values in a interval of length ℓ_i. Let $S = Y_1 + \cdots + Y_n$. Then for every $x \geqslant 0$,*

$$\mathbf{P}(S \geqslant \mathbf{E}[S] + x) \leqslant \exp(-2x^2/L^2),$$

with $L^2 = \ell_1^2 + \cdots + \ell_n^2$.

For f, $g \in \mathcal{F}_0$, we have

$$\mathbf{P}(X_f - X_g > x) = \mathbf{P}\left(\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - g(X_i)) > x\right) \leqslant \exp\left(-\frac{2nx^2}{\|f-g\|_\infty^2}\right),$$

showing that the process $(X_f)_{f\in\mathcal{F}_0}$ is subGaussian with constant $\alpha = 2n$ with respect to the metric $d(f,g) = \|f-g\|_\infty$. Dudley's inequality implies that

$$\mathbf{E}\sup_{f\in\mathcal{F}_0} X_f \leqslant \frac{C}{\sqrt{n}}\int_0^\infty \sqrt{\log N(\mathcal{F}_0, d, \varepsilon)}\,\mathrm{d}\varepsilon.$$

We have $N(\mathcal{F}_0, d, \varepsilon) = 1$ for $\varepsilon \geqslant 1$. For smaller ε, we claim that

$$N(\mathcal{F}_0, d, \varepsilon) \leqslant \left(\frac{C}{\varepsilon}\right)^{C/\varepsilon}. \tag{6.8}$$

It follows that

$$\mathbf{E}\sup_{f\in\mathcal{F}_0} X_f \leqslant \frac{C}{\sqrt{n}}\int_0^1 \frac{\sqrt{\log\varepsilon}}{\sqrt{\varepsilon}}\,\mathrm{d}\varepsilon \leqslant \frac{C'}{\sqrt{n}}.$$

To justify (6.8), consider piece-wise affine functions (check!). □

6.4 VC-dimension

Let Ω any set and $\mathcal{F} \subset \{0,1\}^\Omega$ be a class of functions from Ω to $\{0,1\}$. We say that $\Lambda \subset \Omega$ is *shattered* by $\mathcal{F}$ if any $g : \Lambda \to \{0,1\}$ appears as the restriction to Λ of some $f \in \mathcal{F}$. The *Vapnik–Chervonenkis dimension* of $\mathcal{F}$, denoted by $\mathrm{vc}(\mathcal{F})$, is the largest cardinality of a subset $\Lambda \subset \Omega$ shattered by $\mathcal{F}$.

Here are some examples

1. Let $\Omega = \mathbf{R}$, and $\mathcal{F}$ be the family of indicator functions of segments of $\mathbf{R}$. We have $\mathrm{vc}(\mathcal{F}) = 2$. Indeed, it can checked for example that $\{3,5\}$ is shattered by $\mathcal{F}$. On the other hand, a set $\{a,b,c\}$ with $a < b < c$ cannot be shattered, since no function $f \in \mathcal{F}$ satisfies $f(a) = f(c) = 1$ and $f(b) = 0$.

2. Let $\Omega = \mathbf{R}^2$, and $\mathcal{F}$ be the family of indicator functions of closed half-spaces. Then $\mathrm{vc}(\mathcal{F}) = 3$ (check!).

3. Let $\Omega = \mathbf{R}^2$ and $\mathcal{F}$ be the family of indicator function of convex bodies. Then $\mathrm{vc}(\mathcal{F}) = +\infty$.

Our goal is to prove the following theorem

Theorem 62 (Empirical processes via VC dimension)**.** *Let $\mathcal{F} \subset \{0,1\}^\Omega$, where (Ω, Σ, μ) is a probability space. Let $Z, (Z_i)$ be i.i.d. random variables with law μ. Then*

$$\mathbf{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(Z_i) - \mathbf{E}\, f(Z) \right| \leqslant C \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}.$$

Corollary 63 (Glivenko–Cantelli theorem)**.** *Let (Z_i) be i.i.d. random variables with cumulative distribution function $F(x) = \mathbf{P}(Z_i \leqslant x)$. Consider the empirical distribution function $F_n(x) = \frac{1}{n} \operatorname{card}\{i \in \{1, \cdots, n\} \ : \ Z_i \leqslant x$. Then*

$$\mathbf{E}\, \|F_n - F\|_\infty \leqslant \frac{C}{\sqrt{n}}$$

Proof. Apply Theorem 62 to the family $\{\mathbf{1}_{(-\infty,x]} \ : \ x \in \mathbf{R}\}$, whose VC-dimension equals 2. □

Proof of Theorem 62. We first use a symmetrization argument: if (Z_i') are independent copies of (Z_i), and (ε_i) are independent random signs, then

$$\begin{aligned}
\mathbf{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(Z_i) - \mathbf{E}\, f(Z) \right| &= \mathbf{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(Z_i) - \mathbf{E}\, \frac{1}{n} \sum_{i=1}^n f(Z_i') \right| \\
&\leqslant \mathbf{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(Z_i) - f(Z_i') \right| \\
&= \mathbf{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (f(Z_i) - f(Z_i')) \right| \\
&\leqslant 2\, \mathbf{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(Z_i). \right|
\end{aligned}$$

Define the process $(X_f)_{f \in \mathcal{F}}$ by

$$X_f = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(Z_i).$$

We are going to estimate $\mathbf{E} \sup X_f$ instead of $\mathbf{E} \sup |X_f|$, but this can be easily adapted (check!).

We now work conditionally on the value of (Z_i) (so that the remaining source of randomness comes from the random signs (ε_i)). Conditionally on (Z_i), we have by Hoeffding's inequality

$$\mathbf{P}(X_f - X_g > x) = \mathbf{P}\left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i (f-g)(Z_i) > x \right) \leqslant \exp\left(- \frac{2nx^2}{\frac{1}{n} \sum |(f-g)(Z_i)|^2} \right),$$

which shows that $(X_f)_{f \in \mathcal{F}}$ is subGaussian with constant $2n$ with respect to the (random) distance $d_Z(f,g) = \|f-g\|_{L^2(\mu_Z)}$, where $\mu_Z = \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i}$ is the empirical probability measure associated to $(Z_1, \dots, Z_n)$. We apply Dudley's inequality (conditionally to Z_i) to write

$$\mathbf{E} \sup_{f \in \mathcal{F}} X_f \leqslant \frac{C}{\sqrt{n}} \mathbf{E}_{(Z_i)} \int_0^\infty \sqrt{\log N(\mathcal{F}, d_Z, \varepsilon)} \, \mathrm{d}\varepsilon.$$

It remains to use the following proposition, applied for $\mu = \mu_Z$ to obtain

$$\mathbf{E} \sup_{f \in \mathcal{F}} X_f \leqslant \frac{C}{\sqrt{n}} \mathbf{E}_{(Z_i)} \int_0^\infty \sqrt{\mathrm{vc}(\mathcal{F}) \log(C/\varepsilon)} \, \mathrm{d}\varepsilon \leqslant C \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}. \qquad \square$$

Proposition 64. *Let $\mathcal{F} \subset \{0,1\}^\Omega$, where (Ω, Σ, μ) is a probability space. Then for every $\varepsilon > 0$,*

$$N(\mathcal{F}, L^2(\mu), \varepsilon) \leqslant \left(\frac{C}{\varepsilon} \right)^{C \,\mathrm{vc}(\mathcal{F})}.$$

The proof is based on the following lemmas

Lemma 65. *Let (Ω, Σ, μ) be a probability space, and $\{f_1, \dots, f_N\}$ be an ε-separated set in $L^2(\mu)$. Then there exists a finite subset $\Omega' \subset \Omega$ with $\mathrm{card}(\Omega') \leqslant C\varepsilon^{-4} \log N$ such that $\{f_1, \dots, f_N\}$ is $\varepsilon/2$-separated in $L^2(\nu)$, where ν denotes the uniform probability measure on Ω'.*

Lemma 66 (Sauer–Shelah lemma)**.** *If $\mathcal{F} \subset \{0,1\}^n$ satisfies $\mathrm{vc}(\mathcal{F}) = d$, then*

$$\mathrm{card}(\mathcal{F}) \leqslant \sum_{k=0}^{d} \binom{n}{k} \leqslant \left(\frac{en}{d} \right)^d.$$

Proof of Proposition 64. Using (3.3), there is a subset $P = \{f_1, \dots, f_N\} \subset \mathcal{F}$ with $N = N(\mathcal{F}, L^2(\mu), \varepsilon)$ which is ε-separated in $L^2(\mu)$ norm. Let Ω' be the set produced by applying Lemma 65 to these functions. Let $P' \subset \{0,1\}^{\Omega'}$ be the set of restrictions to Ω' of elements from P. We have $\operatorname{card}(P') = \operatorname{card}(P)$ since P is $\varepsilon/2$-separated in $L^2(\nu)$ (check!).

By the Sauer–Shelah lemma (applied to $P' \subset \{0,1\}^{\Omega'}$), we have, denoting $d = \operatorname{vc}(P')$,

$$N = \operatorname{card}(P') \leqslant \left(\frac{e \operatorname{card} \Omega'}{d} \right)^d \leqslant \left(\frac{C\varepsilon^{-4} \log N}{d} \right)^d$$

and therefore (check!) $N \leqslant (C\varepsilon^{-4})^{2d}$. Finally, it is obvious that $\operatorname{vc}(P') \leqslant \operatorname{vc}(P) \leqslant \operatorname{vc}(\mathcal{F})$. □

Proof of Lemma 65. Choose $\Omega' = \{x_1, \dots, x_n\}$ at random, with (x_i) being i.i.d. of law μ. For $i \neq j$, let $h = (f_i - f_j)^2$. We have

$$\|f_i - f_j\|^2_{L^2(\nu)} - \|f_i - f_j\|^2_{L^2(\mu)} = \frac{1}{n} \sum_{i=1}^{n} h(x_i) - \mathbf{E}\, h(x).$$

Since h is bounded by 1, Hoeffding's inequality applies and yields

$$\mathbf{P} \left(\left| \|f_i - f_j\|^2_{L^2(\nu)} - \|f_i - f_j\|^2_{L^2(\mu)} \right| > x \right) \leqslant 2 \exp(-2nx^2).$$

Since $\|f_i - f_j\|^2_{L^2(\mu)} \geqslant \varepsilon^2$, we have (chose $x = 3\varepsilon^2/4$)

$$\mathbf{P} \left(\|f_i - f_j\|_{L^2(\nu)} > \frac{\varepsilon^2}{4} \right) \leqslant 2 \exp(-cn\varepsilon^4).$$

By the union bound, we obtain that

$$\mathbf{P} \left(\{f_1, \dots, f_N\} \text{ is not } \varepsilon/2\text{-separated in } L^2(\nu) \right) \leqslant 2N^2 \exp(-cn\varepsilon^4)$$

which is less that 1 for $n = C\varepsilon^{-4} \log N$. □

Proof of Lemma 66. We prove a stronger statement: any family $\mathcal{F} \subset \{0,1\}^n$ shatters at least $\operatorname{card}(\mathcal{F})$ subsets of $\{0,1\}$.

We proceed by induction on $\operatorname{card}(\mathcal{F})$. Any $\mathcal{F}$ shatters the empty set. If $\operatorname{card} \mathcal{F} \geqslant 2$, then there is $x \in \{1, \dots, n\}$ and $f_1, f_2 \in \mathcal{F}$ such that $f_1(x) \neq f_2(x)$. Define subfamilies

$$\mathcal{F}_0 = \{f \in \mathcal{F} \ : \ f(x) = 0\},$$
$$\mathcal{F}_1 = \{f \in \mathcal{F} \ : \ f(x) = 1\}.$$

By induction, $\mathcal{F}_0$ (resp. $\mathcal{F}_1$) shatters at least $\operatorname{card}(\mathcal{F}_0)$ (resp. $\operatorname{card}(\mathcal{F}_1)$) subsets of $\{0,1\}^n$. Let S be a subset shattered by $\mathcal{F}_0$ or $\mathcal{F}_1$. Note that S cannot contain x.

- Obviously, S is shattered by $\mathcal{F}$.
- If S is shattered by both $\mathcal{F}_0$ and $\mathcal{F}_1$, then $S \cup \{x\}$ is also shattered by $\mathcal{F}$.

This shows that the number of sets shattered by $\mathcal{F}$ it at least $\mathrm{card}(\mathcal{F}_0)+\mathrm{card}(\mathcal{F}_1) = \mathrm{card}(\mathcal{F})$, as needed. The last inequality in (66) is elementary (check!). $\square$