Asymptotic Convex Geometry
Lecture Notes

Tomasz Tkocz*

These lecture notes were written for the course 21-801 *An introduction to asymptotic convex geometry* that I taught at Carnegie Mellon University in Fall 2018.

*Carnegie Mellon University; ttkocz@math.cmu.edu

Contents

1 Convexity

We shall work in the n-dimensional Euclidean space $\mathbb{R}^n$ equipped with the **standard scalar product** which is defined for two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ as $\langle x, y\rangle = x_1y_1 + \ldots + x_ny_n$, which gives rise to the **standard Euclidean** norm $|x| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + \ldots + x_n^2}$. The (closed) **Euclidean unit ball** is of course defined as $B_2^n = \{x \in \mathbb{R}^n, |x| \le 1\}$ and its boundary is the (Euclidean) **unit sphere** $S^{n-1} = \partial B_2^n = \{x \in \mathbb{R}^n, |x| = 1\}$.

1.1 Sets

For two nonempty subsets A and B of $\mathbb{R}^n$, their **Minkowski sum** is defined as $A + B = \{a + b,\ a \in A, b \in B\}$. The **dilation** of A by a real number t is defined as $tA = \{ta, a \in A\}$. In particular, $-A = \{-a,\ a \in A\}$ and A is called **symmetric** if $-A = A$, that is a point a belongs to A if and only if its symmetric image $-a$ belongs to A. For example, the Minkowski sum of a singleton $\{v\}$ and a set A, abbreviated as $A + v$ is the translate of A by v. The Minkowski sum of the set A and the ball of radius r is the r-enlargement of A, that is the set of points x whose distance to A, $\mathrm{dist}(x, A) = \inf\{|x - a|, a \in A\}$ is at most r, $A_r = A + rB_2^n = \{x \in \mathbb{R}^n, \mathrm{dist}(x, A) \le r\}$. The Minkowski sum of two segments $[a, b]$ and $[c, d]$ is the parallelogram at $a+c$ spanned by $b - a$ and $d - c$, that is $[a, b] + [c, d] = a + c + \{s(b - a) + t(d - c),\ s, t \in [0, 1]\}$ (a **segment** joining two points a and b is of course the set $\{\lambda a + (1 - \lambda)b, \lambda \in [0, 1]\}$.

A subset A of $\mathbb{R}^n$ is called **convex** if along with every two points in the set, the set contains the segment joining them: for every $a, b \in A$ and $\lambda \in [0, 1]$, we have $\lambda a + (1 - \lambda)b \in A$. In other words, A is convex if for every $\lambda \in [0, 1]$, the Minkowski sum $\lambda A + (1 - \lambda)A$ is a subset of A. By induction, A is convex if and only if, for any points $a_1, \ldots, a_k$ in A and weights $\lambda_1, \ldots, \lambda_k \ge 0$, $\sum \lambda_i = 1$, the **convex combination** $\lambda_1 a_1 + \ldots + \lambda_k a_k$ belongs to A. For example, subspaces as well as affine subspaces are convex; particularly, **hyperplanes**, that is co-dimension one affine subspaces, $H = \{x \in \mathbb{R}^n, \langle x, v\rangle = t\}$, $v \in \mathbb{R}^n$, $t \in \mathbb{R}$. Moreover, **half-spaces** $H^- = \{x \in \mathbb{R}^n, \langle x, v\rangle \le t\}$, $H^+ = \{x \in \mathbb{R}^n, \langle x, v\rangle \le t\}$ are convex.

Straight from definition, intersections of convex sets are convex, thus it makes sense to define the smallest convex set containing a given set $A \subset \mathbb{R}^n$ as

$$\mathrm{conv}\, A = \bigcap \{B,\ B \supset A,\ B \text{ convex}\},$$

called its **convex hull**. For instance, the convex hull of the four points $(\pm 1, \pm 1)$ on the plane is the square $[-1, 1]^2$. Plainly,

$$\mathrm{conv}\, A = \left\{ \sum_{i=1}^{k} \lambda_i a_i,\ k \ge 1, a_i \in A, \lambda_i \ge 0, \sum \lambda_i = 1 \right\}$$

(conv A is contained in any convex set containing A, particularly the set on the right is such a set; conversely, if $B \supset A$ for a convex set B, then the set on the right is contained in B, thus it is contained in the intersection of all such sets, which is conv A). This can be compared with the notion of the affine hull,

$$\operatorname{aff}(A) = \left\{ \sum_{i=1}^{k} \lambda_i a_i,\ k \geq 1, a_i \in A, \lambda_i \in \mathbb{R}, \sum \lambda_i = 1 \right\},$$

which is the smallest affine subspace containing A.

The intersection of finitely many closed half-spaces is called a **polyhedral set**, or simply a **polyhedron**. The convex hull of finitely many points is called a **polytope**. In particular, the convex hull of $r+1$ affine independent points is called an r**-simplex**.

A basic theorem in combinatorial geometry due to Carathéodory asserts that points from convex hulls can in fact be expresses as convex combinations of only dimension plus one many points.

1.1 Theorem (Carathéodory)**.** *Let A be a subset of $\mathbb{R}^n$ and let x belong to* conv A*. Then*

$$x = \lambda_1 a_1 + \ldots + \lambda_{n+1} a_{n+1}$$

for some points $a_1, \ldots, a_{n+1}$ from A and nonnegative weights $\lambda_1, \ldots, \lambda_{n+1}$ adding up to 1*.*

Proof. For $y \in \mathbb{R}^n$ and $t \in \mathbb{R}$ by $\left[\begin{smallmatrix} y \\ t \end{smallmatrix}\right]$ we mean the vector in $\mathbb{R}^{n+1}$ whose last component is t and the first n are given by y. Since x belongs to conv A, we can write for some $a_1, \ldots, a_k$ from A and nonnegative $\lambda_1, \ldots, \lambda_k$,

$$\left[\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right] = \sum_{i=1}^{k} \lambda_i \left[\begin{smallmatrix} a_i \\ 1 \end{smallmatrix}\right]$$

(the last equation taking care of $\sum \lambda_i = 1$). Let k be the smallest possible for which this is possible. We can assume that the λ_i used for that are positive. We want to show that $k \leq n+1$. If not, $k > n+2$, the vectors $\left[\begin{smallmatrix} a_1 \\ 1 \end{smallmatrix}\right], \ldots, \left[\begin{smallmatrix} a_k \\ 1 \end{smallmatrix}\right]$ are not linearly independent, thus there are reals $\mu_1, \ldots, \mu_k$, not all zero, such that

$$\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] = \sum_{i=1}^{k} \mu_i \left[\begin{smallmatrix} a_i \\ 1 \end{smallmatrix}\right].$$

Therefore, for every $t \in \mathbb{R}$ we get

$$\left[\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right] = \sum_{i=1}^{k} (\lambda_i + t\mu_i) \left[\begin{smallmatrix} a_i \\ 1 \end{smallmatrix}\right].$$

Notice that the weights $\lambda_i + t\mu_i$ are all positive for $t = 0$, so they all remain positive for small t and there is a choice for t so that (at least) one of the weights becomes zero with the rest remaining positive. This contradicts the minimality of k. $\square$

In particular, Carathéodory theorem says that convex sets can be covered with n-simplices. On the other hand, convex sets are nothing but intersections of half-spaces. To show this, we start with the fact that closed convex sets admit unique closest points, which lies at the heart of convexity.

1.2 Theorem. *For a closed convex set K in $\mathbb{R}^n$ and a point x outside K, there is a unique closest point to x in K (closest in the Euclidean metric).*

Proof. The existence of a closest point follows since K is closed (if $d = \mathrm{dist}(x, K)$, then $d = \mathrm{dist}(x, K \cap RB_2^n)$ for a large $R > 0$, say $R = |x| + d + 1$, consequently there is a sequence of points y_n in $K \cap RB_2^n$ such that $|x - y_n| \to d$ and by compactness we can assume that y_n converges to, say y which is in $K \cap RB_2^n$ and $|x - y| = d$).

The uniqueness of a closest point follows since K is convex and Euclidean balls are round (strictly convex): if y and y' are two different points in K which are closest to x, then the point $\frac{y+y'}{2}$ is in K and is closer to x because by the parallelogram identity,

$$\left|\frac{y-x}{2} + \frac{y'-x}{2}\right|^2 + \left|\frac{y-x}{2} - \frac{y'-x}{2}\right|^2 = 2\left(\left|\frac{y-x}{2}\right|^2 + \left|\frac{y'-x}{2}\right|^2\right) = \mathrm{dist}(x, K)^2$$

and note that the second term on the left $\left|\frac{y-x}{2} - \frac{y'-x}{2}\right|^2 = \left|\frac{y-y'}{2}\right|^2$ is positive, which gives that the first term $\left|\frac{y-x}{2} + \frac{y'-x}{2}\right|^2 = \left|\frac{y+y'}{2} - x\right|^2$ has to be smaller than $\mathrm{dist}(x, K)^2$, that is $\frac{y+y'}{2}$ is closer to x. □

This theorem allows to easily construct separating hyperplanes. A hyperplane $H = \{x \in \mathbb{R}^n,\ \langle x, v\rangle = t\}$ is called a **supporting hyperplane** for a closed convex set K in $\mathbb{R}^n$, if K lies entirely on one side of H, that is K is in either $H^- = \{x \in \mathbb{R}^n,\ \langle x, v\rangle \le t\}$ or $H^+ = \{x \in \mathbb{R}^n,\ \langle x, v\rangle \ge t\}$, and H touches K, that is $H \cap K \neq \varnothing$. Then the set $H \cap K$ of contact points is called a **support set**.

1.3 Theorem. *Let K be a closed and convex set in $\mathbb{R}^n$, let x be a point outside K. Then x can be separated from K by a supporting hyperplane.*

Proof. Let y be the closest point in K to x and let H be the hyperplane which passes through y and is perpendicular to $y - x$. We claim that K lies entirely on the other side of H than x (for if not, there is a closer point in K to x than y – picture). □

1.4 Corollary. *Every closed convex set in $\mathbb{R}^n$ is an intersection of closed half-spaces.*

Proof. For every $x \notin K$, let H_x be the supporting separating hyperplane constrcuted in Theorem 1.3 and say $K \subset H_x^+$. Then clearly, $K \subset \bigcap_{x \in K} H_x^+$, but also $K^c \subset \bigcup_{x \in K} (H_x^+)^c$ because $x \in (H_x^+)^c$, which together proves that $K = \bigcap_{x \in K} H_x^+$. □

By virtue of Theorem 1.3, it makes sense to define the **closest point function**, a sort of projection: for a closed convex set K, let $P_K : \mathbb{R}^n \to \mathbb{R}^n$ be defined by

$$P_K(x) = \text{the closest point in } K \text{ to } x.$$

We remark that this function is 1-Lipschitz.

1.5 Theorem. *Let K be a closed and convex set in $\mathbb{R}^n$. Then the closest point function P_K is 1-Lipschitz (with respect to the Euclidean metric).*

Proof. Suppose that x is not in K. By the construction of separating hyperplanes in the proof of Theorem 1.3, for every $z \in K$, we have $\langle z - P_K(x), x - P_K(x)\rangle \leq 0$. Putting $z = y$, we get $\langle P_K(y) - P_K(x), x - P_K(x)\rangle \leq 0$. If x is in K, $P_K(x) = x$ and this inequality is trivially true. Thus in any case,

$$\langle P_K(y) - P_K(x), x - P_K(x)\rangle \leq 0.$$

Changing the roles of x and y gives

$$\langle P_K(x) - P_K(y), y - P_K(y)\rangle \leq 0.$$

Adding the last two inequalities gives

$$\langle P_K(y) - P_K(x), x - P_K(x) - y + P_K(y)\rangle \leq 0,$$

hence, rearranging and using the Cauchy-Schwarz inequality yields

$$|P_K(y) - P_K(x)|^2 \leq \langle P_K(y) - P_K(x), y - x\rangle \leq |P_K(y) - P_K(x)| \cdot |y - x|,$$

which finishes the proof. □

It is sometimes convenient to take a supporting hyperplane of a convex set at its boundary point. The existence of such hyperplanes follows from a limiting argument.

1.6 Theorem. *Let K be a closed and convex set in $\mathbb{R}^n$ and let x be a point on its boundary. There is a supporting hyperplane for K at x.*

Proof. Since x is on the boundary of K, there is a sequence of points x_n outside K convergent to x. Let $H_{x_n} = \{y \in \mathbb{R}^n,\ \langle y - P_K(x_n), v_n\rangle = 0\}$ be the supporting hyperplanes from Theorem 1.3 and, say $K \subset H_{x_n}^- = \{y \in \mathbb{R}^n,\ \langle y - P_K(x_n), v_n\rangle \leq 0\}$ contain K. We can assume that the vectors v_n are unit, so by compactness we can also assume that that they converge to a unit vector v. Since P_K is continuous (Theorem 1.5), $P_K(x_n) \to P_K(x)$. Let $H = \{y \in \mathbb{R}^n,\ \langle y - x, v\rangle = 0\}$. This is a supporting hyperplane at x because: of course $x \in H$ and if $y \in K$, we know $\langle y - P_K(x_n), v_n\rangle \leq 0$, so in the limit $\langle y - x, v\rangle \leq 0$, which proves that $K \subset H^-$. □

Recall that a support set for K is the set $K \cap H$ for some supporting hyperplane. For polytopes support sets are called **faces**. They can be 0 to $n-1$ dimensional. The $n-1$ dimensional faces are called **facets** and 1 dimensional faces are called **edges**. For polytopes, faces are again polytopes, which we describe in the following theorem.

1.7 Theorem. *Let $P = \text{conv}\{x_i\}_{i=1}^N$ be a polytope in $\mathbb{R}^n$ and let F be its face. Then $F = \text{conv}\{\{x_i\} \cap F\}$. In particular, P has finitely many faces.*

Proof. Let H be the supporting hyperplane associated with the face F, $F = P \cap H$, say $H = \{x \in \mathbb{R}^n,\ \langle x, v\rangle = t\}$ and $H^- = \{x \in \mathbb{R}^n,\ \langle x, v\rangle \le t\} \supset P$. Let k be the index such that $x_1, \ldots, x_k \in F$ and $x_{k+1}, \ldots, x_N \notin F$ (after relabeling the x_i if needed). Take a positive number δ such that $\langle x_l, v\rangle \le t - \delta$ for all $l \ge k+1$. If we take $x \in F$, we write it as $x = \sum_{i=1}^N \lambda_i x_i$, but then

$$t = \langle x, v\rangle = \sum_{i=1}^N \lambda_i \langle x_i, v\rangle \le \sum_{i=1}^k \lambda_i t + \sum_{i=k+1}^N \lambda_i (t - \delta) = t - \delta \sum_{i=k+1}^N \lambda_i$$

and the right hand side is strictly less than t unless the λ_i are zero for all $i > k$ (or the sum is in fact empty). In any case, this shows that $F = \text{conv}\{\{x_i\}_{i=1}^k\}$. □

1.8 Corollary. *Polytopes are polyhedra.*

Proof. For each of finitely many faces of a polytope P, take its supporting hyperplane and take the intersection of the closed half-spaces containing P those hyperplanes determine. The resulting set is P (check this!). □

The generalisation of vertices of polytopes are extremal points of general convex sets. For a closed convex set K in $\mathbb{R}^n$, a point x in K is called **extremal** if $x = \lambda y + (1-\lambda)z$ with $y, z \in K$ and $\lambda \in (0,1)$ implies that $y = z = x$ (in other words, x is not a nontrivial convex combination of other points from K). The set of the extremal points of K is denoted $\text{ext}(K)$. A point x is called **exposed** if $\{x\} = K \cap H$ for some supporting hyperplane H. The set of the exposed points of K is denoted $\text{expo}(K)$. Note that

1) $\text{expo}(K) \subset \text{ext}(K)$ (exposed points are extremal: say x is exposed and lies on the hyperplane $\{\langle y, v\rangle = t\}$, so for every other point z in K we have $\langle z, v\rangle < t$).

2) Closed half-spaces have no extremal points.

3) $\text{expo}(B_2^n) = \text{ext}(B_2^n) = S^{n-1}$.

4) For a stadium shaped convex body $\text{expo}(K) \subsetneq \text{ext}(K)$.

5) Compact convex sets have exposed points (let K be compact and convex, consider a ball B which contains K and has the smallest possible radius; then a tangency point $y \in \partial K \cap \partial B$ is exposed because the supporting hyperplane for B at y is also supporting for K).

6) For a polytope, the exposed and extremal points are the same and they are the vertices of the polytope.

Minkowski's theorem (a finite dimensional version of the Krein-Milman theorem) generalises the last remark to arbitrary compact convex sets.

1.9 Theorem (Minkowski)**.** *Let K be a compact convex set in $\mathbb{R}^n$. If A is a subset of K, then $K = \operatorname{conv} A$ if and only if $A \supset \operatorname{ext}(K)$. In particular, $K = \operatorname{conv}(\operatorname{ext}(K))$.*

Proof. If $K = \operatorname{conv} A$ and there was a point x which is extremal but not in A, then $A \subset K \setminus \{x\}$, but since x is extremal, $K \setminus \{x\}$ is still convex, so $K = \operatorname{conv} A \subset K \setminus \{x\}$, a contradiction.

For the converse, it is enough to show that $K = \operatorname{conv}\operatorname{ext}(K)$. We do it by induction on the dimension. For $n = 1$, K is a closed (bounded) interval and everything is clear. Let $n \geq 2$ and take $x \in K \setminus \operatorname{ext}(K)$. Our goal is to write x as a convex combination of extremal points. We can write x as a convex combination of two boundary points, $\lambda x_1 + (1-\lambda)x_2$ for $x_1, x_2 \in \partial K$, $\lambda \in (0,1)$ (x as not being extremal is in an interval contained in K, so extend the interval until it hits the boundary). Take a supporting hyperplane H at x_1 and consider $K \cap H$. By induction, x_1 can be written as a convex combination of extremal points of $K \cap H$ which are also extremal for K (check!). Similarly for x_2. □

We can now complement Corollary 1.8 and show that bounded polyhedra are polytopes.

1.10 Corollary. *Bounded polyhedra are polytopes.*

Proof. Let P be a bounded polyhedron. In view of Theorem 1.9, we only want to show that P has finitely many extremal points. Let $P = \cap_{i=1}^m H_i^+$ for some closed half-spaces H_i^+ determined by hyperplanes H_i. Let $x \in \operatorname{ext}(P)$, say $x \in H_1 \cap \ldots \cap H_k$ and $x \notin H_{k+1}, \ldots, H_m$. Consider the following subset of P,

$$H_1 \cap \ldots \cap H_k \cap (H_{k+1}^+ \setminus H_{k+1}) \cap \ldots (H_m^+ \setminus H_m).$$

It contains x, it is relatively open, that is it is open in its affine span (as an intersection of open sets). Since x is extremal, this set cannot contain any neighbourhood of x, so this set has to be the singleton $\{x\}$. Since there are only 2^m sets of such form, there are only at most 2^m extremal points of P. □

We also establish the following analogue of Minkowski's theorem for exposed points.

1.11 Theorem. *For a compact convex set K in $\mathbb{R}^n$, we have $K = \overline{\operatorname{conv}\operatorname{expo}(K)}$.*

Proof. Let $L = \overline{\operatorname{conv}\operatorname{expo}(K)}$, which is clearly in K. If there was a point which is in K but not in L, separate it from L by a ball, say $x' + R'B_2^n$ (first do it with a hyperplane and then choose a ball with a big enough radius). Let R be minimal such that $x' + RB_2^n$ contains K. A tangency point y of the ball $x' + RB_2^n$ and K is exposed, but it is not in L. □

We finish by providing a reverse statement to the obvious one $\operatorname{expo}(K) \subset \operatorname{ext}(K)$, mentioned earlier.

1.12 Theorem (Straszewicz). *For a compact convex set K in $\mathbb{R}^n$, we have*

$$\mathrm{ext}(K) \subset \overline{\mathrm{expo}(K)}.$$

Proof. Let $A = \overline{\mathrm{expo}(K)}$. Since for a bounded set S, $\mathrm{conv}\,\overline{S} = \overline{\mathrm{conv}\,S}$ (check!), we have

$$\mathrm{conv}\,A = \mathrm{conv}\,\overline{\mathrm{expo}(K)} = \overline{\mathrm{conv}\,\mathrm{expo}(K)} = K,$$

(the last equality follows from Theorem 1.11). By Theorem 1.9, $A \supset \mathrm{ext}(K)$. □

1.2 Functions

A function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is called **convex** if its **epigraph**,

$$\mathrm{epi}(f) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R},\ f(x) \leq y\}$$

is a convex subset of $\mathbb{R}^{n+1}$. Equivalently, for every $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$,

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

The **domain** of a convex function is the set where it is finite,

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n, f(x) < \infty\}.$$

Note that it is a convex set.

Convex functions are important in optimisation because local minima are global.

Note that the pointwise supremum of a family of convex functions is convex (taking supremum corresponds to intersecting epigraphs). If the epigraph of a convex function is closed, we can view it as an intersection of closed half-spaces. This gives a sometimes useful representation of a convex function as a supremum of affine functions.

1.13 Theorem. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be convex with closed epigraph. Then $f = \sup_\alpha h_\alpha$ for some affine functions h_α.*

By induction, f is convex if and only if for every $x_1, \ldots, x_m \in \mathbb{R}^n$ and nonnegative $\lambda_1, \ldots, \lambda_m$ adding up to one,

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

Jensen's inequality generalises this statement to arbitrary probability measures. A short proof is available thanks to the previous theorem.

1.14 Theorem (Jensen's inequality). *For a probability measure μ on $\mathbb{R}^n$ and a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$, we have*

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

Equivalently, for a random vector X in $\mathbb{R}^n$,

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

Proof. Suppose the epigraph of f is closed (if it is not, an extra argument is needed, but we omit this). With the aid of Theorem 1.13, we have

$$\mathbb{E}f(X) = \mathbb{E}\sup_\alpha h_\alpha(X) \geq \sup_\alpha \mathbb{E}h_\alpha(X) = \sup_\alpha h_\alpha(\mathbb{E}X) = f(\mathbb{E}X),$$

where the last but one equality holds since the h_α are affine. □

Convex functions have good regularity properties. We summarise them in the next two theorems and omit their proofs.

1.15 Theorem. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function. Then f is continuous in the interior of its domain and Lipschitz continuous on any compact subset of that interior.*

1.16 Theorem. *Let A be an open convex subset of $\mathbb{R}^n$ and let $f : A \to \mathbb{R}$. Then*

(i) for $n = 1$: if f is differentiable, then f is convex if and only if f' is nondecreasing; if f is twice differentiable, then f is convex if and only if f'' is nonnegative

(ii) for $n \geq 1$: if f is differentiable, then f is convex if and only if for every x, y in A, we have

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle;$$

if f is twice differentiable, then f is convex if and only if for every x in A, Hess f (x) is positive semi-definite.

1.3 Sets and functions

For a nonempty convex set K in $\mathbb{R}^n$ we define its **support function** $h_K : \mathbb{R}^n \to (-\infty, +\infty]$ as

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

Note several properties

1) h_K is positively homogeneous, that is $h_K(\lambda u) = \lambda h_K(u)$ for every $u \in \mathbb{R}^n$ and $\lambda \geq 0$

2) h_K is convex (as a supremum of linear functions)

3) h_K is finite if and only if K is bounded

4) for a unit vector u, $h_K(u) + h_K(-u)$ is the width of K in direction u

5) $h_{\overline{K}} = h_K$

6) if $K \subset L$, then $h_K \leq h_L$

7) if $h_K \leq h_L$, then $\overline{K} \subset \overline{L}$ (if there was a point x_0 in $\overline{K}$ but not in $\overline{L}$, then separate it from $\overline{L}$ by a hyperplane, say $H = \{x,\ \langle x, v \rangle = t\}$ such that $H^- = \{x, \langle x, v \rangle \leq t\} \supset L$ and then $h_L(v) = \sup_{x \in L} \langle x, v \rangle \leq t < \langle x_0, v \rangle \leq \sup_{x \in \overline{K}} \langle x, v \rangle = h_K(v)$)

8) in particular, closed convex sets are uniquely determined by their support functions

9) $h_{\lambda K} = \lambda h_K$, for $\lambda \geq 0$

10) $h_{-K}(u) = h_K(-u)$ for every vector u

11) $0 \in \overline{K}$ if and only if $h_K \geq 0$ (this is because $\{0\} \subset \overline{K}$ if and only if $0 = h_{\{0\}} \leq h_K$)

12) $h_{K+L} = h_K + h_L$

13) $h_{\operatorname{conv} K_i} = \sup_i h_{K_i}$

For example, for a polytope $P = \operatorname{conv}\{x_i\}_{i=1}^N$, $h_P(x) = \max_{i \leq N} \langle x_i, x \rangle$ (polytopes' support functions are piecewise linear, which is in fact an "if and only if" statement).

Support functions can be characterised by simple conditions: every positively homogeneous, convex function on $\mathbb{R}^n$ with closed epigraph is the support function of a unique closed convex set in $\mathbb{R}^n$. We leave it without proof.

We finish by explaining the name of the support function of a convex set K in $\mathbb{R}^n$. For $u \in \mathbb{R}^n$ consider the hyperplane $H_u = \{x \in \mathbb{R}^n,\ \langle x, u \rangle = h_K(u)\}$. Then $H_u \cap K$ are the points in K attaining the supremum in the definition of $h_K(u)$. If this set is nonempty, then it is a supporting set and H_u is a supporting hyperplane.

1.4 Norms

A function $p : \mathbb{R}^n \to [0, +\infty)$ is a **norm** if it satisfies

1. $p(\lambda x) = |\lambda| p(x)$, $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ (homogeneity)

2. $p(x + y) \leq p(x) + p(y)$, $x, y \in \mathbb{R}^n$ (the triangle inequality)

3. $p(x) = 0$ if and only if $x = 0$.

If p satisfies only 1) and 2), it is called a **semi-norm**. Note also that these two conditions together imply that p is convex.

Let p be a norm on $\mathbb{R}^n$. Define its unit ball $K = \{x \in \mathbb{R}^n,\ p(x) \leq 1\}$. Then K is closed (because p is continuous on $\mathbb{R}^n$). Moreover, K is symmetric by 1), K is convex by the convexity of p and K is bounded thanks to 3). The continuity of p at 0 implies that K contains a small centred Euclidean ball, in particular it has a nonempty interior. In other words, closed unit balls (with respect to norms on $\mathbb{R}^n$) are symmetric compact convex sets with nonempty interior. This and the next theorem saying that the converse is true as well motive the following definition: a **convex body** in $\mathbb{R}^n$ is a compact convex set with nonempty interior.

1.17 Theorem. *Every symmetric convex body in $\mathbb{R}^n$ is the closed unit ball of a norm on $\mathbb{R}^n$.*

Proof. Given a symmetric convex body K in $\mathbb{R}^n$ we define its (so-called) Minkowski functional

$$p_K(x) = \inf\{t > 0,\ x \in tK\}.$$

It is clear that $\{p_K(x) \leq 1\} = K$, so it remains to check that p_K is a norm (exercise). $\square$

An identical argument argument gives a characterisation of unit balls of semi-norms.

1.18 Theorem. *Every symmetric closed convex set in $\mathbb{R}^n$ with nonempty interior is the closed unit ball of a semi-norm on $\mathbb{R}^n$.*

Let us discuss basic examples and properties.

1) for $p > 0$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ define

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

When $p \geq 1$, this is a norm. Its unit ball is denoted as B_p^n,

$$B_p^n = \{x \in \mathbb{R}^n, \|x\|_p \leq 1\}.$$

The space ℓ_p^n is sometimes referred to as the pair $(\mathbb{R}^n, \|\cdot\|_p)$, that is $\mathbb{R}^n$ equipped with the **p-norm**. In particular, B_1^n is the n-dimensional cross-polytope (in $\mathbb{R}^3$: a symmetric piramid), that is

$$B_1^n = \text{conv}\{-e_1, e_1, \ldots, -e_n, e_n\},$$

where as usual e_j is the standard basis vector whose jth component is one and the rest are zero. Moreover,

$$B_\infty^n = [-1, 1]^n$$

is the symmetric cube. Of course, $\|\cdot\|_2$ is just the Euclidean norm and B_2^n is the (closed) centred Euclidean unit ball in $\mathbb{R}^n$.

2) For $1 \leq p \leq q$ we have $B_p^n \subset B_q^n$ and $\|x\|_p \geq \|x\|_q$, $x \in \mathbb{R}^n$.

3) In general, for two symmetric convex bodies K and L in $\mathbb{R}^n$,

$$K \subset L \qquad \text{if and only if} \qquad \|x\|_K \geq \|x\|_L, x \in \mathbb{R}^n$$

where $\|\cdot\|_K$ is the Minkowski functional of K (the norm associated with K whose unit ball is K).

4) $\|x\|_{\lambda K} = \frac{1}{\lambda}\|x\|_K$, $x \in \mathbb{R}^n$, $\lambda > 0$.

5) For instance, $\|x\| = |x_1|$ is a semi-norm whose unit ball is the strip $\{x \in \mathbb{R}^n,\ |x_1| \leq 1\}$. More generally, given $v \in \mathbb{R}^n$, $p(x) = |\langle x, v\rangle|$, $x \in \mathbb{R}^n$ defines a semi-norm whose unit ball is the strip $\{x \in \mathbb{R}^n,\ -1 \leq \langle x, v\rangle \leq 1\}$.

6) Let K be a symmetric convex body. Then by the symmetry of K, the support functional of K, h_K is even, so homogeneous, $h_K(\lambda x) = |\lambda| h_K(x)$, $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$. Recall that h_K is convex, so combined with its homogeneity, h_K satisfies the triangle inequality. Since K is bounded, h_K is finite. Finally, since K has nonempty interior, $h_K(x) = 0$ if and only if $x = 0$. Therefore, h_K is a norm. Its unit ball will be described in the next section.

7) A norm $\|\cdot\|$ on $\mathbb{R}^n$ is called 1-**unconditional** or simply unconditional in a basis $(u_i)_{i=1}^n$ if $\|\sum \varepsilon_i x_i u_i\| = \|\sum x_i u_i\|$ for any choice of signs $\varepsilon_i \in \{-1, 1\}$ and any $x_i \in \mathbb{R}$. If, in addition, $\|\sum x_{\sigma(i)} u_i\| = \|\sum x_i u_i\|$ for any permutation σ of $\{1, \dots, n\}$ and any $x_i \in \mathbb{R}$, the norm is called **1-symmetric**. For instance, the ℓ_p norms are 1-symmetric in the standard basis $(e_i)_{i=1}^n$. For convex bodies, we simplify these notions restricting it just to the standard basis. Thus, a convex body K in $\mathbb{R}^n$ is called **unconditional** if $(\varepsilon_1 x_1, \dots, \varepsilon_n x_n) \in K$ whenever $(x_1, \dots, x_n) \in K$ for any choice of signs $\varepsilon_i \in \{-1, 1\}$ and any $x \in \mathbb{R}^n$. If, in addition, $(x_{\sigma(1)}, \dots, x_{\sigma(n)}) \in K$ whenever $(x_1, \dots, x_n) \in K$ for any permutation σ of $\{1, \dots, n\}$ and any $x \in K$, then K is called **1-symmetric**.

1.5 Duality

For a convex set K in $\mathbb{R}^n$ containing the origin, we define its **polar** by

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

sometimes referred to as the **dual** of K. Equivalently,

$$\begin{aligned} K^\circ &= \{y \in \mathbb{R}^n,\ h_K(y) \leq 1\} \\ &= \{y \in \mathbb{R}^n,\ \forall x \in K\ \langle x, y\rangle \leq \} \\ &= \bigcap_{x \in K} \{y \in \mathbb{R}^n,\ \langle x, y\rangle \leq 1\}, \end{aligned}$$

that is K° is the closed unit ball of the support functional of K (since $0 \in K$, $h_K \geq 0$ and recall that h_K is positively homogeneous and convex, so it is a semi-norm, possibly taking infinite values). The last inequality expresses K° as the intersection of closed half-spaces, so K° is closed and convex.

Let us have a look at some simple examples and properties.

1) $(B_p^n)^\circ = B_q^n$, for $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$ (this follows from Hölder's inequality).

2) The polar of a segment is a strip, for instance $([-1, 1] \times \{0\}^{n-1})^\circ = [-1, 1] \times \mathbb{R}^{n-1}$, and vice versa.

3) $(\mathrm{conv}(K \cup L))^\circ = K^\circ \cap L^\circ$

4) If $K \subset L$, then $K^\circ \supset L^\circ$.

5) $(K^\circ)^\circ \supset K$, with equality for closed convex sets containing the origin.

6) For $A \in GL_n$, $(AK)^\circ = (A^T)^{-1} K^\circ$.

7) If K is a symmetric convex body, then h_K is a norm whose unit ball is K°, so we have $h_K = \|\cdot\|_{K^\circ}$. Since $(K^\circ)^\circ = K$, we get that the support functional of K° is the norm given by K, $h_{K^\circ} = \|\cdot\|_K$. Therefore,

$$\|x\|_K = h_{K^\circ}(x) = \sup_{y \in K^\circ} \langle x, y \rangle,$$

which gives an expression for a norm as a supremum of linear functions. Finally, note that this representation also implies a Cauchy-Schwarz type inequality: for $y \in K^\circ$, we have $\langle x, y \rangle \leq \|x\|_K$, which by homogeneity extends to

$$\langle x, y \rangle \leq \|x\|_K \|y\|_{K^\circ}, \qquad x, y \in \mathbb{R}^n.$$

This notion of duality agrees with the one known from functional analysis: if a norm $\|\cdot\|$ has a unit ball K, then its dual norm $\|\cdot\|'$ (the operator norm on the space of functionals) has the unit ball which is the polar of K°.

1.6 Distances

For two convex bodies K and L in $\mathbb{R}^n$, we define their **Hausdorff distance** by

$$\begin{aligned} \delta^H(K, L) &= \max\{\max_{x \in K} \operatorname{dist}(x, L), \max_{x \in L} \operatorname{dist}(x, K)\} \\ &= \inf\{\delta > 0,\ K \subset L + \delta B_2^n \text{ and } L \subset K + \delta B_2^n\}. \end{aligned}$$

Since $K \subset L + \delta B_2^n$ if and only if $h_K \leq h_L + \delta$ for all unit vectors, the Hausdorff distance $\delta^H(K, L)$ is the smallest number δ such that $h_K \leq h_L + \delta$ and $h_L \leq h_K + \delta$, that is $|h_K - h_L| \leq \delta$, hence

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

(the Hausdorff distance is the supremum distance on the unit sphere for support functions, which also shows that the Hausdorff distance is a metric on the convex bodies).

We also recall the **Banach-Mazur distance** for *symmetric* convex bodies

$$d_{BM}(K, L) = \inf\{t > 0,\ \exists A \in GL_n\ AK \subset L \subset tAK\},$$

which is linearly invariant: $d_{BM}(SK, TL) = d_{BM}(K, L)$, for any $S, T \in GL_n$.

Similarly, the Banach-Mazur distance between two normed spaces $X = (\mathbb{R}^n, \|\cdot\|_K)$ and $Y = (\mathbb{R}^n, \|\cdot\|_L)$ can be defined as

$$d_{BM}(X, Y) = \inf\{t > 0,\ \exists A \in GL_n\ \forall x\ \|x\|_K \leq \|Ax\|_L \leq t\|x\|_K\}.$$

Note that $\log d_{BM}$ satisfies the triangle inequality. A part of asymptotic convex geometry is concerned with questions about (Banach-Mazur) distances to various spaces, for instance what is the dependence on n of $d_{BM}(\ell_1^n, \ell_2^n)$? For any n-dimensional normed space X, how large can $d_{BM}(X, \ell_2^n)$ be? How about $d_{BM}(X, \ell_\infty^n)$?

1.7 Volume

The n-dimensional volume (Lebesgue measure) of a measurable set A in $\mathbb{R}^n$ is denoted by $|A| = \mathrm{vol}_n(A)$. For a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$, $|TA| = |\det T||A|$. In particular, $|tA| = t^n|A|$, $t \geq 0$. If A is in a lower dimensional affine subspace, say a k-dimensional H, then the k-dimensional volume of A (on H) is denoted by $\mathrm{vol}_k(A) = \mathrm{vol}_H(A)$, sometimes also by $|A|$, if it does not lead to any confusion.

Let σ be the normalised (surface) measure on S^{n-1}. It is a probability measure which can be defined using Lebesgue measure on $\mathbb{R}^n$ by

$$\sigma(A) = \frac{|\mathrm{cone}(A)|}{|B_2^n|},$$

for $A \subset S^{n-1}$, where $\mathrm{cone}(A) = \{ta,\ a \in A, t \in [0,1]\}$. It is a unique rotationally invariant probability measure on the sphere, that is $\sigma(UA) = \sigma(A)$, for any orthogonal map $U \in O(n)$ and measurable subset A of the sphere.

Let us recall integration in polar coordinates. For an integrable function $f : \mathbb{R}^n \to \mathbb{R}$ we have

$$\int_{\mathbb{R}^n} f(x)\mathrm{d}x = \int_{S^{n-1}} \int_0^\infty f(r\theta) r^{n-1} |S^{n-1}| \mathrm{d}\sigma(\theta)\mathrm{d}r$$

because (informally) the volume element $\mathrm{d}x$ becomes $|rS^{n-1}|\mathrm{d}\sigma(\theta)\mathrm{d}r$ and the ($n-1$-dimensional) surface measure of the sphere scales like $|rS^{n-1}| = r^{n-1}|S^{n-1}|$. In particular, if we apply this to the indicator function $\mathbf{1}_K$ of a star-shaped set K in $\mathbb{R}^n$ with the radial function $\rho_K(\theta) = \sup\{r \geq 0,\ r\theta \in K\}$, $\theta \in S^{n-1}$, we obtain

$$\begin{aligned} |K| = \int_{\mathbb{R}^n} \mathbf{1}_K &= \int_{S^{n-1}} \int_0^\infty \mathbf{1}_K(r\theta) r^{n-1} |S^{n-1}| \mathrm{d}\sigma(\theta)\mathrm{d}r \\ &= \int_{S^{n-1}} \left(\int_0^{\rho_K(\theta)} r^{n-1}\mathrm{d}r \right) |S^{n-1}| \mathrm{d}\sigma(\theta) \\ &= \frac{|S^{n-1}|}{n} \int_{S^{n-1}} \rho_K(\theta)^n \mathrm{d}\sigma(\theta). \end{aligned}$$

If K is a symmetric convex body, then its radial function can be expressed using its norm,

$$\rho_K(\theta) = \sup\{r \geq 0,\ r\theta \in K\} = \frac{1}{\inf\{\frac{1}{r},\ r\theta \in K\}} = \frac{1}{\inf\{r,\ r\theta \in K\}} = \frac{1}{\|\theta\|_K},$$

thus

$$|K| = \frac{|S^{n-1}|}{n} \int_{S^{n-1}} \|\theta\|_K^{-n} \mathrm{d}\sigma(\theta).$$

In particular, for $K = B_2^n$, we get $|B_2^n| = \frac{|S^{n-1}|}{n}$.

For instance, how to compute the volume of the B_p^n balls? The above formula is not useful. We can do another trick. Using the homogeneity of volume, for a symmetric convex body K in $\mathbb{R}^n$ and $p > 0$, we have

$$\begin{aligned}\int_{\mathbb{R}^n} e^{-\|x\|_K^p} \mathrm{d}x &= \int_{\mathbb{R}^n} \left(\int_{\|x\|_K^p}^{\infty} e^{-t} \mathrm{d}t \right) \mathrm{d}x = \int_{\mathbb{R}^n} \int_0^{\infty} \mathbf{1}_{\{t > \|x\|_K^p\}} e^{-t} \mathrm{d}t \mathrm{d}x \\ &= \int_0^{\infty} e^{-t} \int_{\mathbb{R}^n} \mathbf{1}_{\{x \in \mathbb{R}^n,\ \|x\|_K < t^{1/p}\}} \mathrm{d}x \mathrm{d}t = |K| \int_0^{\infty} t^{n/p} e^{-t} \mathrm{d}t \\ &= |K| \Gamma\left(1 + \frac{n}{p}\right).\end{aligned}$$

In particular,

$$\begin{aligned}\Gamma(1 + n/p)|B_p^n| &= \int_{\mathbb{R}^n} e^{-\|x\|_p^p} \mathrm{d}x = \int_{\mathbb{R}^n} \prod_{i=1}^n e^{-|x_i|^p} \mathrm{d}x = \left(\int_{\mathbb{R}} e^{-|t|^p} \mathrm{d}t \right)^n \\ &= \left(2\Gamma\left(1 + \frac{1}{p}\right) \right)^n.\end{aligned}$$

Setting $p = 2$ gives us a formula for the volume of the unit Euclidean ball

$$|B_2^n| = 2^n \frac{\Gamma(3/2)^n}{\Gamma(1 + n/2)} = \frac{\sqrt{\pi}^n}{\Gamma(1 + n/2)}. \tag{1.1}$$

By Stirling's formula,

$$|B_2^n| = (1 + o(1)) \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{2\pi e}{n}}^n.$$

This gives us that the radius r_n of the Euclidean ball with volume one,

$$r_n = \frac{1}{\sqrt{\pi}} \Gamma(1 + n/2)^{1/n} = (1 + o(1)) \sqrt{\frac{n}{2\pi e}}.$$

We also get $|B_\infty^n| = 2^n$ and $|B_1^n| = \frac{2^n}{n!}$. Using our previous formula for volume, we have

$$2^n = |B_\infty^n| = |B_2^n| \int_{S^{n-1}} \rho_{B_\infty^n}(\theta)^n \mathrm{d}\sigma(\theta)$$

which means that the radial function of the cube on average equals

$$\rho_{B_\infty^n} \approx 2|B_2^n|^{-1/n} \approx \sqrt{\frac{2n}{\pi e}}.$$

Similarly,

$$\rho_{B_1^n} \approx \frac{2}{n!^{1/n}} |B_2^n|^{-1/n} \approx \sqrt{\frac{2e}{\pi n}}.$$

1.8 Ellipsoids

An **ellipsoid** $\mathcal{E}$ in $\mathbb{R}^n$ is a set of the form

$$\mathcal{E} = \left\{ x \in \mathbb{R}^n,\ \sum_{i=1}^n \frac{\langle x, v_i \rangle^2}{\alpha_i^2} \le 1 \right\},$$

where $\{v_i\}_{i=1}^n$ is an orthonormal basis of $\mathbb{R}^n$ and the α_i are positive numbers. Since $\big\langle x, (\sum \alpha_i^{-2} v_i v_i^T)x \big\rangle = \sum \alpha_i^{-2} \langle x, v_i\rangle^2$, we can also write

$$\mathcal{E} = \{x \in \mathbb{R}^n,\ \langle x, Ax\rangle \leq 1\}\,,$$

where $A = \sum \alpha_i^{-2} v_i v_i^T = V \mathrm{diag}(\alpha_i^{-2}) V^T$ with V being the orthogonal matrix whose columns are the vectors v_i. Note that any positive semi-definite matrix is of this form.

The vectors v_i are the directions of the axes of $\mathcal{E}$ and the α_i are the lengths of the axes. Let T be the linear map on $\mathbb{R}^n$ sending v_i to $\alpha_i v_i$. Then, $\mathcal{E} = TB_2^n$. In particular, we find that the volume is

$$|\mathcal{E}| = (\alpha_1 \cdot \ldots \cdot \alpha_n)|B_2^n| = (\det A)^{-1/2}|B_2^n|. \tag{1.2}$$

The norm $\|\cdot\|_{\mathcal{E}}$ can be expressed explicitly because $\|x\|_{\mathcal{E}} = \inf\{t > 0,\ x \in t\mathcal{E}\} = \inf\{t > 0,\ \langle x, Ax\rangle \leq t^2\}$, so

$$\|x\|_{\mathcal{E}} = \sqrt{\langle x, Ax\rangle} = \sqrt{\sum \frac{\langle x, v_i\rangle^2}{\alpha_i^2}}. \tag{1.3}$$

Any linear image AB_2^n of the unit Euclidean ball is an ellipsoid (possibly lower-dimensional). To see that, consider the singular value decomposition $A = VDU$ with D being a nonnegative diagonal matrix and U, V orthogonal. Of course $UB_2^n = B_2^n$, so $AB_2^n = VDB_2^n$. DB_2^n is an ellipsoid with the axes along the standard basis and the lengths given the diagonal elements of D, so VDB_2^n is the ellipsoid with the axes along the columns of V of such lengths.

2 Log-concavity

2.1 Brunn-Minkowski inequality

Brunn discovered the following concavity property of the volume: for a convex set K in $\mathbb{R}^n$ the function $f(t) = \mathrm{vol}_{n-1}(K \cap (t\theta + \theta^\perp)$, $t \in \mathbb{R}$, of the volumes of the sections of K along a direction $\theta \in S^{n-1}$ is $\frac{1}{n-1}$ concave on its support, that is $f(t)^{\frac{1}{n-1}}$ is concave on its support. Minkowski turned this into a powerful tool.

2.1 Theorem (Brunn-Minkowski inequality)**.** *For nonempty compact sets A, B in $\mathbb{R}^n$ we have*

$$|A + B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}.$$

There are many different proofs. We shall deduce the Brunn-Minkowski inequality from a more general result for functions, the functional inequality due to Prékopa and Leindler. Before that, let us point out several remarks.

2.2 Remark. Thanks to the inner regularity of Lebesgue measure (that is, the Lebesgue measure of a measurable set is the supremum of the Lebesgue measure of its compact subsets), the Brunn-Minkowski inequality extends to arbitrary nonempty measurable sets A and B such that $A + B$ is also measurable: for such sets, let K and L be compact subsets of A and B respectively and then $A + B$ contains $K + L$, so $|A + B|^{1/n} \geq |K + L|^{1/n} \geq |K|^{1/n} + |L|^{1/n}$ and taking the supremum over K and L yields $|A + B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}$.

2.3 Remark. The proof of the Brunn-Minkowski inequality in dimension one is easy. Let A and B be two nonempty compact subsets of $\mathbb{R}$. Thanks to the translation invariance of Lebesgue measure, we can assume that the furthest most right point of A and the furthest left point of B are at the origin. Then the Minkowski sum $A + B$ contains $A \cup B$ whose measure is $|A| + |B|$ because $A \cap B = \{0\}$, so $|A + B| \geq |A \cup B| = |A| + |B|$.

2.4 Remark. To obtain Brunn's concavity principle for the volume of sections of a convex set K in $\mathbb{R}^n$ along a direction $\theta \in S^{n-1}$, define $K_t = \{x \in \theta^\perp,\ x + t\theta \in K\}$, $t \in \mathbb{R}$ and let $f(t)$ be the $n-1$-dimensional volume (on $\theta^\perp$) of K_t. Take $\lambda \in [0, 1]$, s, t in the support of f and set $A = \lambda K_s$ and $B = (1 - \lambda)K_t$. By convexity, $K_{\lambda s + (1-\lambda)t}$ contains $\lambda K_s + (1 - \lambda)K_t = A + B$, thus

$$\begin{aligned} f(\lambda s + (1 - \lambda)t)^{\frac{1}{n-1}} \geq |A + B|^{\frac{1}{n-1}} \geq |A|^{\frac{1}{n-1}} + |B|^{\frac{1}{n-1}} &= \lambda |K_s|^{\frac{1}{n-1}} + (1 - \lambda)|K_t|^{\frac{1}{n-1}} \\ &= \lambda f(s)^{\frac{1}{n-1}} + (1 - \lambda) f(t)^{\frac{1}{n-1}}, \end{aligned}$$

which shows that f is $\frac{1}{n-1}$-concave on its support.

2.5 Remark. The Brunn-Minkowski inequality gives an effortless proof of the isoperimetric inequality.

2.6 Theorem. *For a compact set A in $\mathbb{R}^n$ take a Euclidean ball B with the same volume as A. Then for every $\varepsilon > 0$,*

$$|A + \varepsilon B| \geq |B + \varepsilon B|.$$

In particular, $|\partial A| \geq |\partial B|$.

Proof. By the Brunn-Minkowski inequality and the scaling properties of volume,

$$|A + \varepsilon B|^{1/n} \geq |A|^{1/n} + |\varepsilon B|^{1/n} = |B|^{1/n} + \varepsilon |B|^{1/n} = |(1+\varepsilon)B|^{1/n} = |B + \varepsilon B|^{1/n}.$$

Since, $|\partial A| = c \liminf_{\varepsilon \to 0+} \frac{|A+\varepsilon B| - |A|}{\varepsilon}$, where c is a scaling constant which depends only on the volume of A, the second part follows. □

2.7 Remark. By the AM-GM inequality and homogeneity, the Brunn-Minkowski inequality applied to λA and $(1-\lambda)B$ gives $|\lambda A + (1-\lambda)B|^{1/n} \geq \lambda |A|^{1/n} + (1-\lambda)|B|^{1/n} \geq |A|^{\lambda/n}|B|^{(1-\lambda)/n}$, that is:
for compact sets A, B in $\mathbb{R}^n$ and $\lambda \in [0,1]$,

$$|\lambda A + (1-\lambda)B| \geq |A|^{\lambda}|B|^{1-\lambda}. \tag{2.1}$$

In fact, this dimension free statement is equivalent to the Brunn-Minkowski inequality: apply it to the sets $A/|A|^{1/n}$, $B/|B|^{1/n}$ and $\lambda = \frac{|A|^{1/n}}{|A|^{1/n}+|B|^{1/n}}$ to get

$$\begin{aligned}\frac{|A+B|^{1/n}}{|A|^{1/n}+|B|^{1/n}} = \left|\frac{A+B}{|A|^{1/n}+|B|^{1/n}}\right|^{1/n} &= \left|\lambda \frac{A}{|A|^{1/n}} + (1-\lambda)\frac{B}{|B|^{1/n}}\right|^{1/n} \\ &\geq \left|\frac{A}{|A|^{1/n}}\right|^{\lambda/n} \left|\frac{B}{|B|^{1/n}}\right|^{(1-\lambda)/n} = 1.\end{aligned}$$

2.2 Log-concave measures

A Borel measure μ on $\mathbb{R}^n$ is called **log-concave** if for every Borel sets A, B in $\mathbb{R}^n$ and $\lambda \in [0,1]$, we have

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^{\lambda}\mu(B)^{1-\lambda}.$$

The dimension free version (2.1) of the Brunn-Minkowski inequality says that Lebesgue measure is log-concave. Another crucial example of a log-concave measure is the uniform measure on a convex body K in $\mathbb{R}^n$, that is

$$\mu_K(A) = \frac{|A \cap K|}{|K|}, \qquad A \subset \mathbb{R}^n.$$

The reason being that thanks to the convexity of K, $(\lambda A + (1-\lambda)B) \cap K$ contains $\lambda(A \cap K) + (1-\lambda)(B \cap K)$, thus the log-concavity of Lebesgue measure gives

$$\begin{aligned}\mu_K(\lambda A + (1-\lambda)B) &= \frac{1}{|K|}|(\lambda A + (1-\lambda)B) \cap K| \\ &\geq \frac{1}{|K|}|\lambda(A \cap K) + (1-\lambda)(B \cap K)| \\ &\geq \frac{1}{|K|}|A \cap K|^{\lambda}|B \cap K|^{1-\lambda} = \mu_K(A)^{\lambda}\mu_K(B)^{1-\lambda}.\end{aligned}$$

Note that the density of μ_K (with respect to Lebesgue measure), $f(x) = \frac{1}{|K|}\mathbf{1}_K(x)$, plainly satisfies the pointwise inequality

$$f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}, \qquad x, y \in \mathbb{R}^n, \lambda \in [0,1]. \tag{2.2}$$

In other words, the function $\psi : \mathbb{R}^n \to (-\infty, +\infty]$, $\psi = -\log f = \begin{cases} -\log|K|, & x \in K \\ +\infty, & x \notin K \end{cases}$ is convex.

We say that a function $f : \mathbb{R}^n \to [0, +\infty)$ is **log-concave** if it satisfies (2.2), that is $f = e^{-\psi}$ for some convex function $\psi : \mathbb{R}^n \to (-\infty, +\infty]$.

Summarising, for the two examples of log-concave measures we looked at: Lebesgue measure as well as the uniform measure on a convex body, their densities are log-concave functions. As we shall see in the next two sections, this is not accidental. First we need to discuss the Prékopa-Leindler inequality and, incidentally, finish the proof of the Brunn-Minkowski inequality.

2.3 Prékopa-Leindler inequality

2.8 Theorem (Prékopa-Leindler inequality)**.** *Let* $\lambda \in [0,1]$. *For measurable functions* $f, g, h : \mathbb{R}^n \to [0, +\infty)$ *such that*

$$h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}, \qquad x, y \in \mathbb{R}^n, \tag{2.3}$$

we have

$$\int_{\mathbb{R}^n} h \geq \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda}. \tag{2.4}$$

2.9 Remark. For compact sets A, B in $\mathbb{R}^n$ and $\lambda \in [0,1]$, consider $f = \mathbf{1}_A$, $g = \mathbf{1}_B$ and $h = \mathbf{1}_{\lambda A + (1-\lambda)B}$. Then clearly these functions satisfy the assumption of the Prékopa-Leindler inequality, $h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}$ for all $x, y \in \mathbb{R}^n$. Indeed, if the right hand side is 0, there is nothing to show. Otherwise, $x \in A$ and $y \in B$, so $\lambda x + (1-\lambda)y \in \lambda A + (1-\lambda)B$, so the left hand side is 1, equal to the right hand side. Since $\int f = |A|$, $\int g = |B|$ and $\int h = |\lambda A + (1-\lambda)B|$, the Prékopa-Leinder inequality thus implies (2.1), the dimension free version of the Brunn-Minkowski inequality (equivalent to Theorem 2.1, see Remark 2.7).

Proof of Theorem 2.8. First we prove the theorem in dimension one, that is for $n = 1$ and then, by an inductive argument, we will obtain the theorem for every n.

Let f, g, h be nonnegative measurable functions on $\mathbb{R}$ satisfying (2.3). Without loss of generality, we can assume that f and g are bounded (if not, consider $f_M = \min\{f, M\}$ and $g_M = \min\{g, M\}$ which still satisfy the assumption and the conclusion will carry over to f and g by the monotone convergence theorem). Moreover, we can assume that

$\|f\|_\infty = \|g\|_\infty = 1$ (otherwise, consider $f/\|f\|_\infty$, $g/\|g\|_\infty$ and $h/(\|f\|_\infty^\lambda \|g\|_\infty^{1-\lambda})$). Then,

$$\int_{\mathbb{R}} f(x)\mathrm{d}x = \int_{\mathbb{R}}\int_0^{f(x)} \mathrm{d}t\mathrm{d}x = \int_{\mathbb{R}}\int_0^{\|f\|_\infty} \mathbf{1}_{\{f(x)>t\}}\mathrm{d}t\mathrm{d}x = \int_0^1 |\{x\in\mathbb{R},\ f(x)>t\}|\mathrm{d}t$$

and similarly for g. Fix $t \in (0,1)$ and consider the sets $A = \{x \in \mathbb{R},\ f(x) > t\}$ and $B = \{x \in \mathbb{R},\ g(x) > t\}$. Note they are nonempty because $\|f\|_\infty = 1 = \|g\|_\infty$. Note also that by (2.3), we have $\lambda A + (1-\lambda)B \subset \{x \in \mathbb{R},\ h(x) > t\}$. By the Brunn-Minowski inequality in dimension one proved in Remark 2.3,

$$\lambda|A| + (1-\lambda)|B| = |\lambda A| + |(1-\lambda)B| \le |\lambda A + (1-\lambda)B| \le |\{x\in\mathbb{R},\ h(x)>t\}|$$

(see also Remark 2.2 to go about a possible lack of compactness of A and B)). Thus, by the AM-GM inequality,

$$\begin{aligned}\left(\int_{\mathbb{R}} f\right)^\lambda \left(\int_{\mathbb{R}} g\right)^{1-\lambda} \le \lambda\int_{\mathbb{R}} f + (1-\lambda)\int_{\mathbb{R}} g &= \int_0^1 (\lambda|A| + (1-\lambda)|B|)\mathrm{d}t \\ &\le \int_0^1 |\{x\in\mathbb{R},\ h(x)>t\}|\mathrm{d}t \\ &\le \int_0^\infty |\{x\in\mathbb{R},\ h(x)>t\}|\mathrm{d}t = \int_{\mathbb{R}} h,\end{aligned}$$

which finishes the proof for $n = 1$.

Let $n > 1$ and suppose the theorem holds in any dimension less than n. For $t \in \mathbb{R}$ define three functions on $\mathbb{R}^{n-1}$ with parameter t being the restrictions of f, g and h:

$$f_t(x') = f(t,x'), \quad g_t(x') = g(t,x') \quad \text{and} \quad h_t(x') = h(t,x'), \qquad x' \in \mathbb{R}^{n-1}.$$

Fix $s, t \in \mathbb{R}$. From (2.3), for every $x', y' \in \mathbb{R}^{n-1}$,

$$h_{\lambda s+(1-\lambda)t}(\lambda x' + (1-\lambda)y') \ge f_s(x')^\lambda g_t(y')^{1-\lambda}.$$

Thus, by the inductive assumption,

$$\int_{\mathbb{R}^{n-1}} h_{\lambda s+(1-\lambda)t} \ge \left(\int_{\mathbb{R}^{n-1}} f_s\right)^\lambda \left(\int_{\mathbb{R}^{n-1}} g_t\right)^\lambda.$$

This however says that the three functions F, G and H on $\mathbb{R}$ defined as

$$F(t) = \int_{\mathbb{R}^{n-1}} f_t, \quad G(t) = \int_{\mathbb{R}^{n-1}} g_t \quad \text{and} \quad H(t) = \int_{\mathbb{R}^{n-1}} h_t, \qquad t \in \mathbb{R}$$

satisfy $H(\lambda s + (1-\lambda)t) \ge F(s)^\lambda G(t)^{1-\lambda}$ for every $s, t \in \mathbb{R}$. Therefore, by the $n = 1$ case of the theorem proved earlier,

$$\int_{\mathbb{R}} H \ge \left(\int_{\mathbb{R}} F\right)^\lambda \left(\int_{\mathbb{R}} G\right)^{1-\lambda},$$

which is exactly $\int_{\mathbb{R}^n} h \ge \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda}$, as needed. □

2.4 Basic properties of log-concave functions

Let us point out several important consequences of the Prékopa-Leindler inequality. Firstly, log-concave functions form a class of functions closed with respect to taking marginals and convolutions.

2.10 Corollary. *If a function $f : \mathbb{R}^m \times \mathbb{R}^n \to [0, +\infty)$ is log-concave, then the function*

$$h(x) = \int_{\mathbb{R}^n} f(x, y)\mathrm{d}y, \qquad x \in \mathbb{R}^m$$

is also log-concave.

Proof. This follows directly from the Prékopa-Leindler inequality: to see that $h(\lambda x_1 + (1-\lambda)x_2) \geq h(x_1)^\lambda h(x_2)^{1-\lambda}$, it suffices to consider three functions $F(y) = f(x_1, y)$, $G(y) = f(x_2, y)$, $H(y) = f(\lambda x_1 + (1-\lambda)x_2, y)$, $y \in \mathbb{R}^n$. □

2.11 Corollary. *If functions $f, g : \mathbb{R}^n \to [0, +\infty)$ are log-concave, then their convolution $f \star g$ is also log-concave.*

Proof. Apply Corollary 2.10 to $\mathbb{R}^n \times \mathbb{R}^n \ni (x, y) \mapsto f(y)g(x-y)$ which is log-concave. □

Secondly, measures with log-concave densities are log-concave.

2.12 Corollary. *If $f : \mathbb{R}^n \to [0, +\infty)$ is log-concave, then the Borel measure μ with density f, defined for Borel subsets A in $\mathbb{R}^n$ by*

$$\mu(A) = \int_A f,$$

is log-concave.

Proof. Given two Borel sets A, B in $\mathbb{R}^n$ and $\lambda \in [0, 1]$, apply the Prékopa-Leindler inequality to the three functions $f\mathbf{1}_A$, $f\mathbf{1}_B$ and $f\mathbf{1}_{\lambda A+(1-\lambda)B}$ to see that $\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}$. □

2.13 Remark. The same argument shows that for a log-concave function $f : \mathbb{R}^n \to [0, +\infty)$ which is supported on a lower-dimensional affine subspace H of $\mathbb{R}^n$ (f is zero outside H), the measure μ on $\mathbb{R}^n$ defined by

$$\mu(A) = \int_{A\cap H} f(x)\mathrm{dvol}_H(x), \qquad A \subset \mathbb{R}^n$$

is log-concave.

This says that absolutely continuous measures (with respect to Lebesgue measure on a possibly lower dimensional affine subspace) whose densities are log-concave are log-concave measures. In particular, uniform measures on convex sets are log-concave, as we have already observed in the case of convex bodies. Note that this includes point masses, a.k.a. Dirac delta measures. Moreover, Gaussian measures are log-concave.

Another examples of log-concave measures are the products of exponential or Gamma measures.

The converse to the above corollary is also true, which is a deep result of Borell (we omit its proof).

2.14 Theorem (Borell)**.** *If a finite inner-regular measure (a finite measure approximable from below by compact sets) μ on $\mathbb{R}^n$ is log-concave, then there is an affine subspace H of $\mathbb{R}^n$ and a log-concave function $f : \mathbb{R}^n \to [0, +\infty)$ which is zero outside H and such that*

$$\mu(A) = \int_{A \cap H} f(x) \mathrm{dvol}_H(x).$$

In particular, the support of μ, that is the set $\{x \in \mathbb{R}^n,\ \mu(x + rB_2^n) > 0$ for every $r > 0\}$ is contained in H.

Together with Corollary 2.12, Borell's theorem provides the charactetrisation saying that (finite) log-concave measures are absolutely continuous measures (on the affine span of their support) with log-concave densities.

Let us recapitulate this discussion in the probabilistic language. A random vector X in $\mathbb{R}^n$ is called **log-concave** if its distribution

$$\mu(A) = \mathbb{P}\left(X \in A\right), \qquad A \subset \mathbb{R}^n,$$

is a log-concave measure. As we saw, by the Prékopa-Leindler and Borell's theorems, a random vector is log-concave if and only if it is supported on some affine subspace, continuous on it, with a log-concave density.

For instance, a random vector X in $\mathbb{R}^3$ uniformly distributed on the square $[-1, 1]^2 \times \{0\}$ is log-concave. Even though X is not continuous (as a vector in $\mathbb{R}^3$), it has a density on $\mathbb{R}^2 \times \{0\}$ which is uniform, $\frac{1}{4}\mathbf{1}_{[-1,1]^2}$.

2.15 Corollary. *If X is a log-concave random vector in $\mathbb{R}^n$, then its marginals are also log-concave. Even more, for any affine map $A : \mathbb{R}^n \to \mathbb{R}^m$, the vector AX is also log-concave.*

Proof. For Borel subsets U, V of $\mathbb{R}^m$ and $\lambda \in [0, 1]$, we have

$$\begin{aligned}
\mathbb{P}\left(AX \in \lambda U + (1-\lambda)V\right) &= \mathbb{P}\left(X \in A^{-1}(\lambda U + (1-\lambda)V)\right) \\
&= \mathbb{P}\left(X \in \lambda A^{-1}U + (1-\lambda)A^{-1}V\right) \\
&\geq \mathbb{P}\left(X \in A^{-1}U\right)^{\lambda} \mathbb{P}\left(X \in A^{-1}V\right)^{1-\lambda} \\
&= \mathbb{P}\left(AX \in U\right)^{\lambda} \mathbb{P}\left(AX \in V\right)^{1-\lambda}.
\end{aligned}$$

□

2.16 Corollary. *If X is a log-concave random vector in $\mathbb{R}^n$ and Y is an independent log-concave random vector in $\mathbb{R}^m$, then (X, Y) is a log-concave random vector in $\mathbb{R}^{n+m}$.*

Proof. By Borell's theorem, X has a density f on an affine subspace F of $\mathbb{R}^n$ and Y has a density g on an affine subspace H of $\mathbb{R}^m$. Then (X,Y) has the product density $f(x)g(y)$ on $\{(x,y), x\in F, y\in H\} = F\times H$, which is a log-concave function. By the Prékopa-Leindler inequality (as in Remark 2.13), (X,Y) is log-concave. □

2.17 Corollary. *If X and Y are independent log-concave random vectors on $\mathbb{R}^n$, then $X+Y$ is also log-concave.*

Proof. Since $X+Y$ is the linear image of (X,Y), the assertion follows directly from Corollaries 2.15 and 2.16. □

We finish with a useful fact for log-concave random variables (random vectors in $\mathbb{R}$) saying that their PDFs and tails are also log-concave.

2.18 Corollary. *If X is a log-concave random variable, then the functions $\mathbb{R}\ni t\mapsto \mathbb{P}(X\le t)$ and $\mathbb{R}\ni t\mapsto \mathbb{P}(X>t)$ are log-concave.*

Proof. It follows immediately from the definition (note that $\{X\le t\} = \{X\in(-\infty,t]\}$ and $\{X>t\} = \{X\in(t,\infty)\}$). □

2.5 Further properties of log-concave functions

Log-concave functions decay at least exponentially fast; particularly, they have good integrability properties, for instance, their moments are finite (we skip the standard proof).

2.19 Theorem. *Let $f:\mathbb{R}^n\to[0,+\infty)$ be an integrable log-concave function. Then there are positive constants A,α such that $f(x)\le Ae^{-\alpha|x|}$, for all $x\in\mathbb{R}^n$. In particular, for every $p>-n$, $\int_{\mathbb{R}^n}|x|^p f(x)\mathrm{d}x<\infty$.*

Centred log-concave functions have their value at the origin comparable to the maximum.

2.20 Theorem. *Let $f:\mathbb{R}^n\to[0,+\infty)$ be a centred log-concave function, that is $\int_{\mathbb{R}^n} xf(x)\mathrm{d}x = 0$. Then,*

$$f(0)\le\|f\|_\infty\le e^n f(0).$$

Proof. Without loss of generality we can assume that $\int f = 1$. By log-concavity, for every $x,y\in\mathbb{R}^n$ and $\lambda\in[0,1]$, we have

$$f(\lambda x+(1-\lambda)y)\ge f(x)^\lambda f(y)^{1-\lambda}.$$

First integrating over y and then taking the supremum over x yields

$$(1-\lambda)^{-n}\ge\|f\|_\infty^\lambda\int f^{1-\lambda}.$$

We have equality at $\lambda = 0$, thus differentiating at $\lambda = 0$ gives

$$n \geq \log \|f\|_\infty - \int f \log f.$$

Rearranging and using Jensen's inequality (for the concave function $\log f$ and probability measure with density f) finishes the proof

$$\log\left(e^{-n}\|f\|_\infty\right) \leq \int f(x)\log f(x)\mathrm{d}x \leq \log f\left(\int xf(x)\mathrm{d}x\right) = \log f(0).$$

□

The next result is the monotonicity of some sort of moments of log-concave functions on the half-line. This gives the optimal moment comparison for log-concave random variables, which can be viewed as a reverse Hölder-type inequality.

2.21 Theorem. *Let $f : [0,\infty) \to [0,\infty)$ be a log-concave function with $f(0) > 0$. Then the function*

$$(0,\infty) \ni p \mapsto \left(\frac{1}{\Gamma(p)}\frac{1}{f(0)}\int_0^\infty f(x)x^{p-1}\mathrm{d}x\right)^{1/p}$$

is nonincreasing.

2.22 Corollary. *For a nonnegative random variable X with a log-concave tail,*

$$\left(\mathbb{E}X^q\right)^{1/q} \leq \frac{\Gamma(q+1)^{1/q}}{\Gamma(p+1)^{1/p}}\left(\mathbb{E}X^q\right)^{1/q}, \qquad 0 < p < q.$$

Equality holds for $X \sim Exp(1)$ (one-sided standard exponential).

Proof. Apply Theorem 2.21 to $f(x) = \mathbb{P}(X \geq x)$ and note that $f(0) = 1$ as well as

$$\left(\frac{1}{\Gamma(p+1)}\mathbb{E}X^p\right)^{1/p} = \left(\frac{1}{\Gamma(p)p}\frac{1}{f(0)}\int_0^\infty px^{p-1}f(x)\mathrm{d}x\right)^{1/p}.$$

□

In view of Corollary 2.18, the above moment comparison holds for nonnegative log-concave random variables and consequently, for symmetric random variables.

Proof of Theorem 2.21. Without loss of generality, $f(0) = 1$. Fix $0 < p < q$. The idea is to compare any log-concave function f to the extremal one (exponential) with the same value at 0. Take $\alpha > 0$ such that

$$\int_0^\infty e^{-\alpha x}x^{p-1}\mathrm{d}x = \int_0^\infty f(x)x^{p-1}\mathrm{d}x.$$

By looking at the logs, $f(x)$ and $e^{-\alpha x}$ intersect at some point, say $x = c$ and $f(x) - e^{-\alpha x}$ is nonnegative on $[0,c]$ and nonpositive on $[c,\infty)$. Thus,

$$\int_0^\infty f(x)x^{q-1} - \int_0^\infty e^{-\alpha x}x^{q-1} = \int_0^\infty x^{q-p}\left[f(x) - e^{-\alpha x}\right]x^{p-1}.$$

On $[0,c]$, $x^{q-p} \le c^{q-p}$ and $f(x) - e^{-\alpha x}$ is nonnegative, whereas on $[c,\infty)$, $x^{q-p} \ge c^{q-p}$ but $f(x) - e^{-\alpha x}$ is nonpositive, thus

$$\int_0^\infty f(x)x^{q-1} - \int_0^\infty e^{-\alpha x}x^{q-1} \\ \le c^{q-p}\left(\int_0^c \left[f(x) - e^{-\alpha x}\right]x^{p-1} + \int_c^\infty \left[f(x) - e^{-\alpha x}\right]x^{p-1}\right) = 0.$$

Computing the integrals with $e^{-\alpha x}$ finishes the proof. ☐

There is also a useful reverse monotonicity result which holds for all functions.

2.23 Theorem. *Let $f : [0,\infty) \to [0,\infty)$ be a measurable function. Then the function*

$$(0,\infty) \ni p \mapsto \left(\frac{p}{\|f\|_\infty}\int_0^\infty f(x)x^{p-1}\mathrm{d}x\right)^{1/p}$$

is nondecreasing.

Proof. Without loss of generality, $\|f\|_\infty = 1$. Let $F(p) = \left(p\int_0^\infty f(x)x^{p-1}\mathrm{d}x\right)^{1/p}$. Fix $0 < p < q$. For any $a > 0$,

$$\begin{aligned}\frac{F(q)^q}{q} = \int_0^\infty f(x)x^{q-1} &= \int_0^a f(x)x^{q-1} + \int_a^\infty x^{q-p}f(x)x^{p-1} \\ &\ge \int_0^a f(x)x^{q-1} + a^{q-p}\int_a^\infty f(x)x^{p-1} \\ &= \int_0^a f(x)x^{q-1} + a^{q-p}\int_0^\infty f(x)x^{p-1} - a^{q-p}\int_0^a f(x)x^{p-1} \\ &= a^{q-p}\frac{F(p)^p}{p} + a^{q-1}\int_0^a f(x)\left[\left(\frac{x}{a}\right)^{q-1} - \left(\frac{x}{a}\right)^{p-1}\right].\end{aligned}$$

Note that $(x/a)^{q-1} - (x/a)^{p-1} \le 0$ on $[0,a]$. Thus bounding in the last integral f by 1 (its supremum) yields

$$\frac{F(q)^q}{q} \ge a^{q-p}\frac{F(p)^p}{p} + a^q\left(\frac{1}{q} - \frac{1}{p}\right).$$

Putting $a = F(p)$ finishes the proof. ☐

We finish with a corollary saying that the variance of a log-concave function is comparable to the square of the reciprocal of its value at its centre (which is in turn comparable to its maximal value, as we already know).

2.24 Corollary. *Let $f : \mathbb{R} \to [0,\infty)$ be a centred log-concave function. Then,*

$$\frac{1}{12e^2} \le \frac{f(0)^2\int x^2 f(x)\mathrm{d}x}{\left(\int f\right)^3} \le 2.$$

Proof. The right inequality follows from Theorem 2.21. The left inequality follows from Theorems 2.20 and 2.23. ☐

2.6 Ball's inequality

We conclude this chapter with a functional inequality due to Ball, which is of a similar flavour as the Prékopa-Leindler inequality. It is also a good excuse to present a different proof-technique of such inequalities, based on transport of mass. This technique can be applied to give another proof of the one dimensional case of the Prékopa-Leindler inequality.

2.25 Theorem (Ball's inequality)**.** *Let $p > 0$ and $\lambda \in [0,1]$. If three integrable functions $u, v, w : (0,+\infty) \to [0,+\infty)$ satisfy for all positive r, s,*

$$w\left((\lambda r^{-1/p} + (1-\lambda)s^{-1/p})^{-p}\right) \geq u(r)^{\frac{\lambda r^{-1/p}}{\lambda r^{-1/p}+(1-\lambda)s^{-1/p}}} v(s)^{\frac{(1-\lambda)s^{-1/p}}{\lambda r^{-1/p}+(1-\lambda)s^{-1/p}}}, \qquad (2.5)$$

then

$$\int_0^\infty w \geq \left(\lambda\left(\int_0^\infty u\right)^{-1/p} + (1-\lambda)\left(\int_0^\infty v\right)^{-1/p}\right)^{-p}. \qquad (2.6)$$

Proof. Without loss of generality, we can modify u, v, w as follows: first we can assume that u, v are bounded and supported in a compact subset of $(0,+\infty)$ (otherwise consider $u_M = \min\{u, M\}\mathbf{1}_{[1/M,M]}$ which monotonely converges to u and similar modifications for v and w). Now, we can assume that u, v, w are strictly positive on $[1/M, M]$ (otherwise consider $u+\varepsilon$, $v+\varepsilon$ and $w+\varepsilon'$). Moreover, we can also assume that u and v are continuous (otherwise approximate from below u and v by monotone sequences of continuous functions). Having established the assertion for such modified functions, it yields the assertion for initial functions by Lebesgue's monotone convergence theorem.

Define two strictly increasing functions $\alpha, \beta : [0,1] \to (0,\infty)$ by

$$\int_0^{\alpha(t)} u = t\int_0^\infty u \qquad \text{and} \qquad \int_0^{\beta(t)} v = t\int_0^\infty v.$$

They are differentiable and satisfy

$$\alpha'(t)u(\alpha(t)) = \int u \qquad \text{and} \qquad \beta'(t)v(\beta(t)) = \int v.$$

Set

$$\gamma(t) = \left(\lambda\alpha(t)^{-1/p} + (1-\lambda)\beta(t)^{-1/p}\right)^{-p}, \qquad t > 0.$$

This is a strictly increasing differentiable function mapping $(0,1)$ onto $(0,\infty)$, thus by the change of variables we have

$$\int_0^\infty w = \int_0^1 w(\gamma(t))\gamma'(t)\mathrm{d}t.$$

Our goal is to estimate γ' from below. We have

$$\begin{aligned}\gamma'(t) &= -p\gamma(t)^{-p-1}\left(-\frac{1}{p}\lambda\alpha(t)^{-1/p-1}\alpha'(t) - \frac{1}{p}(1-\lambda)\beta(t)^{-1/p-1}\beta'(t)\right)\\ &= \gamma^{1+1/p}\left(\frac{\lambda\alpha'}{\alpha^{1+1/p}} + \frac{(1-\lambda)\beta'}{\beta^{1+1/p}}\right)\\ &= \lambda\left(\frac{\gamma}{\alpha}\right)^{1+1/p}\frac{\int u}{u(\alpha)} + (1-\lambda)\left(\frac{\gamma}{\beta}\right)^{1+1/p}\frac{\int v}{v(\beta)}.\end{aligned}$$

Note that

$$\left(\frac{\gamma}{\alpha}\right)^{1/p} = \frac{\alpha^{-1/p}}{\lambda\alpha^{-1/p} + (1-\lambda)\beta^{-1/p}},$$

thus setting

$$\theta = \frac{\lambda\alpha^{-1/p}}{\lambda\alpha^{-1/p} + (1-\lambda)\beta^{-1/p}},$$

the expression for γ' becomes

$$\gamma' = \theta\left(\frac{\theta}{\lambda}\right)^p \frac{\int u}{u(\alpha)} + (1-\theta)\left(\frac{1-\theta}{1-\lambda}\right)^p \frac{\int v}{v(\beta)}$$

and by the AM-GM inequality we get

$$\gamma' \geq \left[\left(\frac{\theta}{\lambda}\right)^p \frac{\int u}{u(\alpha)}\right]^\theta \left[\left(\frac{1-\theta}{1-\lambda}\right)^p \frac{\int v}{v(\beta)}\right]^{1-\theta}.$$

This, assumption (2.5) and the inequality $x^\theta y^{1-\theta} \geq (\theta x^{-1/p} + (1-\theta)y^{-1/p})^{-p}$ allow to finish the proof,

$$\begin{aligned}
\int w = \int_0^1 w(\gamma)\gamma' &\geq \int_0^1 u(\alpha)^\theta v(\beta)^{1-\theta} \left[\left(\frac{\theta}{\lambda}\right)^p \frac{\int u}{u(\alpha)}\right]^\theta \left[\left(\frac{1-\theta}{1-\lambda}\right)^p \frac{\int v}{v(\beta)}\right]^{1-\theta} \\
&\geq \int_0^1 \left[\theta\frac{\lambda}{\theta}\left(\int u\right)^{-1/p} + (1-\theta)\frac{1-\lambda}{1-\theta}\left(\int v\right)^{-1/p}\right]^{-p} \\
&= \left[\lambda\left(\int u\right)^{-1/p} + (1-\lambda)\left(\int v\right)^{-1/p}\right]^{-p}.
\end{aligned}$$

□

3 Concentration

Concentration of measure is a powerful and influential idea in high dimensional analysis and probability. In essence, this is a phenomenon when in a certain probability space with a metric structure, decent size sets rapidly become of almost full measure, when enlarging them just a little bit. We shall analyse in detail three basic examples of the sphere, Gaussian space and discrete cube. We shall also establish concentration for log-concave measures and as applications show moment comparison inequalities.

3.1 Sphere

We treat the unit Euclidean sphere S^{n-1} in $\mathbb{R}^n$ as a probability space with its Haar measure σ and a metric space with simply the Euclidean distance (see the appendix). For a subset A of S^{n-1} and $t \geq 0$, we define its **t-enlargement** A_t by

$$A_t = \{x \in S^{n-1},\ \mathrm{dist}(x, A) \leq t\}.$$

Note that for $t \geq 2$, A_t becomes the whole sphere. The concentration of measure phenomenon on the sphere is expressed in the following result.

3.1 Theorem. *For a Borel subset A of the unit Euclidean sphere S^{n-1} with measure at least one-half, $\sigma(A) \geq 1/2$, we have for positive t,*

$$\sigma(A_t) \geq 1 - 2e^{-nt^2/4}.$$

Proof. We can assume that $t < 2$. Let B be the complement of the t-enlargement A_t of A, $B = S^{n-1} \setminus A_t$. For $x \in A$ and $y \in B$, we have $|x - y| \geq t$, so

$$\left|\frac{x+y}{2}\right| = \sqrt{1 - \left(\frac{|x-y|}{2}\right)^2} \leq \sqrt{1 - \frac{t^2}{4}} \leq 1 - \frac{t^2}{8}.$$

Let $\tilde{A}$ be the part in B_2^n of the cone built on A, $\tilde{A} = \{\alpha x,\ \alpha \in [0,1], x \in A\}$, so that $\sigma(A) = |\tilde{A}|/|B_2^n|$; similarly for B and $\tilde{B}$. Consider $\tilde{x} \in \tilde{A}$ and $\tilde{y} \in \tilde{B}$, say $\tilde{x} = \alpha x$ and $\tilde{y} = \beta y$, for some $\alpha, \beta \in [0,1]$ and $x \in A$, $y \in B$. If, say $\alpha \leq \beta$, we have

$$\begin{aligned}\left|\frac{\tilde{x}+\tilde{y}}{2}\right| = \left|\frac{\alpha x + \beta y}{2}\right| = \beta\left|\frac{\frac{\alpha}{\beta}x + y}{2}\right| &= \beta\left|\frac{\alpha}{\beta}\frac{x+y}{2} + \left(1 - \frac{\alpha}{\beta}\right)\frac{y}{2}\right| \\ &\leq \left|\frac{\alpha}{\beta}\frac{x+y}{2} + \left(1 - \frac{\alpha}{\beta}\right)\frac{y}{2}\right| \\ &\leq \frac{\alpha}{\beta}\left|\frac{x+y}{2}\right| + \left(1 - \frac{\alpha}{\beta}\right)\left|\frac{y}{2}\right|.\end{aligned}$$

Since $\left|\frac{x+y}{2}\right| \leq 1 - \frac{t^2}{8}$ and $\left|\frac{y}{2}\right| \leq \frac{1}{2} \leq 1 - \frac{t^2}{8}$, we get

$$\left|\frac{\tilde{x}+\tilde{y}}{2}\right| \leq 1 - \frac{t^2}{8},$$

thus

$$\frac{\tilde{A}+\tilde{B}}{2} \subset \left(1-\frac{t^2}{8}\right) B_2^n.$$

By the Brunn-Minkowski inequality,

$$\left(1-\frac{t^2}{8}\right)^n |B_2^n| \geq \left|\frac{\tilde{A}+\tilde{B}}{2}\right| \geq \sqrt{|\tilde{A}|\cdot|\tilde{B}|} = |B_2^n|\sqrt{\sigma(A)\sigma(B)}.$$

Using, $\sigma(A) \geq \frac{1}{2}$, $\sigma(B) = 1-\sigma(A_t)$, $1-\frac{t^2}{8} \leq e^{-t^2/8}$ and rearranging finishes the proof. □

Concentration also means that Lipschitz functions are essentially constant; their values concentrate around their median as well as mean and the two are comparable.

Recall that a **median** of a random variable X, denoted $\mathrm{Med}(X)$ is any number m such that $\mathbb{P}(X \geq m) \geq \frac{1}{2}$ and $\mathbb{P}(X \leq m) \geq \frac{1}{2}$.

3.2 Corollary. *Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function. Then for $t > 0$,*

$$\sigma\{f - \mathrm{Med}(f) > t\} \leq 2e^{-nt^2/4} \qquad \textit{and} \qquad \sigma\{f - \mathrm{Med}(f) < -t\} \leq 2e^{-nt^2/4}.$$

In particular,

$$\sigma\{|f - \mathrm{Med}(f)| > t\} \leq 4e^{-nt^2/4}.$$

Moreover,

$$\left|\mathrm{Med}(f) - \int_{S^{n-1}} f \mathrm{d}\sigma\right| \leq \frac{8}{\sqrt{n}}$$

and

$$\sigma\left\{f - \int_{S^{n-1}} f\mathrm{d}\sigma > t\right\} \leq e^{16}e^{-\frac{nt^2}{16}} \quad \textit{and} \quad \sigma\left\{f - \int_{S^{n-1}} f\mathrm{d}\sigma < -t\right\} \leq e^{16}e^{-\frac{nt^2}{16}}.$$

Proof. Let $A = \{f \leq \mathrm{Med}(f)\}$. By the definition of a median, $\sigma(A) \geq \frac{1}{2}$. Since f is 1-Lipschitz, for $t > 0$,

$$A_t \subset \{f \leq \mathrm{Med}(f) + t\}.$$

Indeed, if $y \in A_t$, say $y = x+z$ with $x \in A$ and $|z| \leq t$, then $f(y) = f(x)+f(y)-f(x) \leq f(x) + |y-x| \leq \mathrm{Med}(f) + t)$. Therefore, $A_t^c \supset \{f > \mathrm{Med}(f) + t\}$ and we get

$$\sigma\{f > \mathrm{Med}(f) + t\} \leq \sigma(A_t^c) \leq 2e^{-nt^2/4}.$$

The estimate for the lower tail follows similarly by considering $A = \{f \geq \mathrm{Med}(f)\}$ (or taking $-f$ in what we just proved).

Moreover,

$$\begin{aligned}\left|\mathrm{Med}(f) - \int_{S^{n-1}} f\mathrm{d}\sigma\right| = \left|\int_{S^{n-1}} (\mathrm{Med}(f) - f)\mathrm{d}\sigma\right| &\leq \int_{S^{n-1}} |\mathrm{Med}(f) - f|\,\mathrm{d}\sigma \\ &= \int_0^\infty \sigma\{|f - \mathrm{Med}(f)| > t\}\,\mathrm{d}t \\ &\leq \int_0^\infty 4e^{-nt^2/4}\mathrm{d}t = \frac{4\sqrt{\pi}}{\sqrt{n}} < \frac{8}{\sqrt{n}}.\end{aligned}$$

Thus, for $t > \frac{16}{\sqrt{n}}$,

$$\sigma\left\{f > \int_{S^{n-1}} f\mathrm{d}\sigma + t\right\} \le \sigma\left\{f > \mathrm{Med}(f) - \frac{8}{\sqrt{n}} + t\right\} \le \sigma\left\{f > \mathrm{Med}(f) + \frac{t}{2}\right\}$$
$$\le 2e^{-nt^2/16}.$$

For $t \le \frac{16}{\sqrt{n}}$, $e^{-nt^2/16} \ge e^{-16}$, so trivially, $\sigma\left\{f > \int_{S^{n-1}} f\mathrm{d}\sigma + t\right\} \le 1 \le e^{16}e^{-nt^2/16}$. □

3.3 Remark. In Corollary 3.2, we deduced the concentration for Lipschitz functions from the concentration for sets. It is also possible to go the other way around: having a statement about concentration for Lipschitz functions and applying it to the distance function to a set which is 1-Lipschitz gives the concentration for sets.

Having seen what concentration of measure is about on the concrete example of the sphere, let us say a few words about concentration in an abstract setting. If we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that (Ω, d) is also a metric space, we define enlargements of measurable sets in the usual way: $A_t = \{x \in \Omega,\ d(x, A) \le t\}$. We say that $(\Omega, \mathbb{P}, d)$ satisfies α**-concentration** with a (decay) function $\alpha : [0,\infty) \to [0,\infty)$ such that $\alpha \underset{t\to\infty}{\longrightarrow} 0$ if for every measurable set A with $\mathbb{P}(A) \ge \frac{1}{2}$, we have $\mathbb{P}(A_t) \ge 1 - \alpha(t)$, $t > 0$. In particular, when $\alpha(t) \approx e^{-t^2}$, it is the so-called **Gaussian concentration** and when $\alpha(t) \approx e^{-t}$ – **exponential concentration**. In Theorem 3.1 we proved that the sphere satisfies Gaussian concentration.

3.2 Gaussian space

We consider $\mathbb{R}^n$ as a probability space equipped with the standard Gaussian measure γ_n and as a metric space with the Euclidean distance. Recall that γ_n has the product density $\frac{1}{\sqrt{2\pi}}e^{-|x|^2/2}$. This setting is usually referred to as **Gaussian space**. We show that Gaussian space satisfies Gaussian concentration.

3.4 Theorem. *For a Borel subset A of $\mathbb{R}^n$,*

$$\int_{\mathbb{R}^n} e^{\frac{1}{4}\,\mathrm{dist}(x,A)^2}\mathrm{d}\gamma_n(x) \le \frac{1}{\gamma_n(A)}.$$

In particular, if $\gamma_n(A) \ge \frac{1}{2}$, then for $t > 0$,

$$\gamma_n(A_t) \ge 1 - 2e^{-t^2/4}.$$

Proof. The second part follows from the main statement in one line,

$$e^{t^2/4}\gamma_n(A_t^c) \le \int_{A_t^c} e^{\mathrm{dist}(x,A)^2/4}\mathrm{d}\gamma_n(x) \le \frac{1}{\gamma_n(A)} \le 2.$$

To prove the first part, fix A, let p_n be the density of γ_n and consider three functions,

$$f(x) = e^{\mathrm{dist}(x,A)^2/4}p_n(x), \qquad g(x) = \mathbf{1}_A(x)p_n(x), \qquad h(x) = p_n(x).$$

For $x \in \mathbb{R}^n$, $y \in A$,

$$f(x)g(y) = e^{\mathrm{dist}(x,A)^2/4} p_n(x)p_n(y) = \frac{1}{(2\pi)^n} e^{\frac{1}{4}\,\mathrm{dist}(x,A)^2 - \frac{1}{2}|x|^2 - \frac{1}{2}|y|^2}$$
$$\leq \frac{1}{(2\pi)^n} e^{\frac{1}{4}|x-y|^2 - \frac{1}{2}|x|^2 - \frac{1}{2}|y|^2}$$
$$= \frac{1}{(2\pi)^n} e^{-\frac{1}{4}|x+y|^2} = p_n\left(\frac{x+y}{2}\right)^2 = h\left(\frac{x+y}{2}\right)^2,$$

so by the Prékopa-Leindler inequality, $\int f \cdot \int g \leq (\int h)^2$, which is the desired statement.

□

As for the sphere, we also obtain the concentration for Lipschitz functions. The proof is identical.

3.5 Corollary. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a 1-Lipschitz function. Then for $t > 0$,*

$$\gamma_n\{f - \mathrm{Med}(f) > t\} \leq 2e^{-t^2/4} \qquad \textit{and} \qquad \gamma_n\{f - \mathrm{Med}(f) < -t\} \leq 2e^{-t^2/4}.$$

In particular,

$$\gamma_n\{|f - \mathrm{Med}(f)| > t\} \leq 4e^{-t^2/4}.$$

Moreover,

$$\left|\mathrm{Med}(f) - \int_{\mathbb{R}^n} f \mathrm{d}\gamma_n\right| \leq 8$$

and

$$\gamma_n\left\{f - \int_{\mathbb{R}^n} f \mathrm{d}\gamma_n > t\right\} \leq e^{16} e^{-\frac{t^2}{16}} \quad \textit{and} \quad \gamma_n\left\{f - \int_{\mathbb{R}^n} f \mathrm{d}\gamma_n < -t\right\} \leq e^{16} e^{-\frac{t^2}{16}}.$$

3.6 Remark. By a different, direct argument based on the rotational invariance of the Gaussian measure, it is possible to obtain the concentration of Lipschitz functions around the mean with better constants, namely

$$\gamma_n\left\{\left|f - \int_{\mathbb{R}^n} f \mathrm{d}\gamma_n\right| > t\right\} \leq \frac{4}{\pi} e^{-\frac{2}{\pi^2}t^2}.$$

3.3 Discrete cube

Consider the **discrete cube** $\Omega_n = \{-1, 1\}^n$ with the uniform probability measure $\mathbb{P}_n$. In other words, for any subset A of Ω_n, $\mathbb{P}_n(A) = \mathbb{P}(\varepsilon \in A) = \frac{|A|}{2^n}$, where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ is a random vector of independent (symmetric) **random signs**, that is the ε_i are i.i.d. taking the values ± 1 with probability $\frac{1}{2}$. We shall write $\mathbb{P}(A) = \mathbb{P}_n(A)$.

The discrete cube, being a subset of $\mathbb{R}^n$, is naturally equipped with the Euclidean distance. There is also the notion of the **Hamming distance** which counts at how many coordinates points differ,

$$d_H(x, y) = |\{i \leq n,\ x_i \neq y_i\}|, \qquad x, y \in \Omega_n.$$

This distance is of course very natural, especially from a combinatorial point of view. Plainly, for $x, y \in \Omega_n$,

$$|x-y|^2 = 4d_H(x,y).$$

Note also that

$$\langle x, y\rangle = n - 2|\{i \le n,\ x_i \neq y_i\}| = n - 2d_H(x,y),$$

or to put it differently, since $n = \langle x, x\rangle$,

$$\langle x, x-y\rangle = 2d_H(x,y).$$

The main theorem we shall prove is the so-called infimum convolution on the discrete cube for convex functions. Then we shall deduce Talagrand's inequality on the discrete cube. Finally, we shall present corollaries which lead to concentration.

3.7 Theorem. *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then,*

$$\mathbb{E}_\varepsilon e^{\inf\{\varphi(x)+\frac{1}{8}|\varepsilon-x|^2,\ x\in\mathbb{R}^n\}}\mathbb{E}_\varepsilon e^{-\varphi(\varepsilon)} \le 1. \tag{3.1}$$

3.8 Corollary (Talagrand's inequality)**.** *Let A be a subset of $\{-1,1\}^n$. Then,*

$$\mathbb{E}_\varepsilon e^{\frac{1}{8}\operatorname{dist}(\varepsilon,\operatorname{conv}(A))^2}\mathbb{P}(A) \le 1.$$

In particular,

$$\mathbb{P}\left((\operatorname{conv} A)_t^c\right)\cdot\mathbb{P}(A) \le e^{-t^2/8} \tag{3.2}$$

Proof. Consider the convex function $\varphi(x) = \begin{cases} 0, & x \in \operatorname{conv}(A) \\ +\infty, & x \in \mathbb{R}^n \setminus \operatorname{conv}(A) \end{cases}$. Since,

$$\mathbb{E}e^{-\varphi(\varepsilon)} = \mathbb{E}\mathbf{1}_{\operatorname{conv}(A)}(\varepsilon) = \mathbb{P}(\operatorname{conv}(A)) \ge \mathbb{P}(A)$$

and

$$\inf_{x\in\mathbb{R}^n}\left\{\varphi(x) + \frac{1}{8}|\varepsilon-x|^2\right\} = \inf_{x\in\operatorname{conv}(A)}\frac{1}{8}|\varepsilon-x|^2 = \frac{1}{8}\operatorname{dist}(\varepsilon,\operatorname{conv} A)^2,$$

applying (3.1) gives the first result. The second part follows by Chebyshev's inequality,

$$\frac{1}{\mathbb{P}(A)} \ge \mathbb{E}e^{\frac{1}{8}\operatorname{dist}(\varepsilon,\operatorname{conv} A)^2} \ge \mathbb{E}e^{\frac{1}{8}t^2}\mathbf{1}_{\{\operatorname{dist}(\varepsilon,\operatorname{conv} A)\ge t\}} \ge e^{t^2/8}\mathbb{P}(\operatorname{dist}(\varepsilon, A) \ge t).$$

□

We also obtain concentration for Lipschitz functions which are additionally *convex*.

3.9 Corollary. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and 1-Lipschitz function (with respect to the Euclidean distance). Then for $t > 0$,*

$$\mathbb{P}\left(|f(\varepsilon) - \operatorname{Med}(f(\varepsilon))| \ge t\right) \le 4e^{-t^2/8}.$$

Proof. Let $M = \mathrm{Med}(f(\varepsilon))$ and consider $A = \{x \in \{-1,1\}^n,\ f(x) \le M\}$. We have $\mathbb{P}(A) \ge \frac{1}{2}$. Since f is convex, A is convex and $\mathrm{conv}\, A = A$. Since f is 1-Lipschitz, $A_t \subset \{f \le M + t\}$. Thus, by (3.2),

$$\mathbb{P}(f > M + t) \le \mathbb{P}(A_t^c) \le \frac{1}{\mathbb{P}(A)} e^{-t^2/8} \le 2e^{-t^2/8}.$$

To control the lower tail, we cannot simply apply the upper tail to $-f$ (because it is not convex). Consider $B = \{x \in \{-1,1\}^n, f(x) \le M - t\}$. As before, B is convex and $B_s \subset \{f \le M - t + s\}$. Thus,

$$\mathbb{P}(f \le M - t)\, \mathbb{P}(f > M - t + s) \le \mathbb{P}(B)\, \mathbb{P}(B_s^c) \le e^{-s^2/2}.$$

Letting $s \nearrow t$ and using $\mathbb{P}(f \ge M) \ge \frac{1}{2}$ gives the result. □

3.10 Remark. As for the sphere and Gaussian space, we get from Corollary 3.9 concentration around the mean: there are universal constants $c, C > 0$ such that for a convex and 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{P}(|f(\varepsilon) - \mathbb{E}f(\varepsilon)| \ge t) \le Ce^{-ct^2}, \qquad t > 0.$$

Note that martingale methods (e.g. Azuma's inequality) yields

$$\mathbb{P}(|f(\varepsilon) - \mathbb{E}f(\varepsilon)| \ge t) \le Ce^{-ct^2/n}, \qquad t > 0,$$

for every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$. Therefore convexity through Talagrand's inequality allows to significantly improve the exponent by removing $\frac{1}{n}$. Moreover, this improved concentration result really is specific to convex functions, as shown by the following example due to Talagrand: let $A = \{\{-1,1\}^n,\ \sum x_i \le 0\}$ and define $f(x) = \inf\{|x-y|, y \in A\}$, which is 1-Lipschitz. Then the median of $f(\varepsilon)$ is 0, but by the central limit theorem, $\mathbb{P}\left(f(\varepsilon) > cn^{1/4}\right) \ge c$ for some absolute constant $c > 0$.

We can easily derive concentration for sets with enlargements in the Hamming distance. The point being however that Talagrand's inequality is stronger than what we really need to get such concentration (the latter can be obtained by classical martingale methods as well).

3.11 Corollary. *Let A be a subset of $\{-1,1\}^n$ with $\mathbb{P}(A) \ge \frac{1}{2}$. Then,*

$$\mathbb{P}(d_H(\varepsilon, A) > t) \le 2e^{-\frac{t^2}{/}2n}.$$

Proof. For every $x \in \{-1,1\}^n$ and $y \in A$,

$$\langle x, x - y\rangle = 2d_H(x,y) \ge 2d_H(x,A).$$

Since the left hand side is linear in y and the right hand side does not depend on y, the same inequality is true for all $y \in \mathrm{conv}(A)$. By the Cauchy-Schwarz inequailty,

$$\sqrt{n}|x - y| \ge 2d_H(x,A), \qquad y \in \mathrm{conv}(A),$$

which means that $d(x, \operatorname{conv} A) \geq \frac{2}{\sqrt{n}} d_H(x,A)$, so

$$\{x \in \{-1,1\}^n,\ d_H(x,A) \geq t\} \subset \left\{x \in \{-1,1\}^n,\ d(x, \operatorname{conv} A) > \frac{2}{\sqrt{n}} t\right\}$$

and Corollary 3.8 finishes the argument. □

Proof of Theorem 3.7. The strangely looking left hand side allows a natural inductive proof on the dimension n (in fancy terms, the inequality tensorises, meaning that it proves itself provided it is proved for $n = 1$ – similarly to the Prékopa-Leindler inequality).

Suppose Theorem 3.7 holds for $n = 1$. We first show how to inductively deduce it for any n. To this end, let $n \geq 1$ and suppose that the theorem holds for n and let $\varphi : \mathbb{R}^{n+1} \to \mathbb{R}$ be a convex function. For a vector x in $\mathbb{R}^{n+1}$ we shall write $x = (x_1, x')$ with $x_1 \in \mathbb{R}$ being its first coordinate and $x' \in \mathbb{R}^n$. Consider the function $\psi : \mathbb{R} \to \mathbb{R}$,

$$\psi(x_1) = \log \mathbb{E}_{\varepsilon'} e^{\inf_{x' \in \mathbb{R}^n} \{\varphi(x_1,x') + \frac{1}{8}|\varepsilon' - x'|^2\}}.$$

By the convexity of φ and $|\cdot|^2$ as well as Hölder's inequality, we check that ψ is also convex: for $x_1, y_1 \in \mathbb{R}$ and $\lambda \in [0,1]$, we have

$$\begin{aligned}
&\psi(\lambda x_1 + (1-\lambda)y_1) \\
&\quad = \log \mathbb{E}_{\varepsilon'} e^{\inf_{x',y' \in \mathbb{R}^n} \left\{\varphi(\lambda x_1 + (1-\lambda)y_1, \lambda x' + (1-\lambda)y') + \frac{1}{8}|\lambda\varepsilon' + (1-\lambda)\varepsilon' - \lambda x' - (1-\lambda)y'|\right\}} \\
&\quad \leq \log \mathbb{E}_{\varepsilon'} e^{\inf_{x',y'} \left\{\lambda[\varphi(x_1,x') + \frac{1}{8}|\varepsilon'-x'|^2] + (1-\lambda)[\varphi(y_1,y') + \frac{1}{8}|\varepsilon'-y'|^2]\right\}} \\
&\quad = \log \mathbb{E}_{\varepsilon'} \left[e^{\lambda \inf_{x'} \left\{\varphi(x_1,x') + \frac{1}{8}|\varepsilon'-x'|^2\right\}} e^{(1-\lambda)\inf_{y'} \left\{[\varphi(y_1,y') + \frac{1}{8}|\varepsilon'-y'|^2]\right\}}\right] \\
&\quad \leq \lambda \log \mathbb{E}_{\varepsilon'} e^{\inf_{x'} \left\{\varphi(x_1,x') + \frac{1}{8}|\varepsilon'-x'|^2\right\}} + (1-\lambda) \log \mathbb{E}_{\varepsilon'} e^{\inf_{y'} \left\{[\varphi(y_1,y') + \frac{1}{8}|\varepsilon'-y'|^2]\right\}} \\
&\quad = \lambda\psi(x_1) + (1-\lambda)\psi(y_1).
\end{aligned}$$

Applying (3.1) to ψ yields

$$\underbrace{\mathbb{E}_{\varepsilon_1} e^{\inf_{x_1} \{\psi(x_1) + \frac{1}{8}|\varepsilon_1 - x_1|^2\}}}_{A} \cdot \underbrace{\mathbb{E}_{\varepsilon_1} e^{-\psi(\varepsilon_1)}}_{B} \leq 1.$$

To lower bound A, we rewrite it, plug in the definition of ψ and simplify,

$$\begin{aligned}
A &= \mathbb{E}_{\varepsilon_1} \inf_{x_1} \left[e^{\psi(x_1)} e^{\frac{1}{8}|\varepsilon_1 - x_1|^2}\right] \\
&= \mathbb{E}_{\varepsilon_1} \inf_{x_1} \left[\mathbb{E}_{\varepsilon'} e^{\inf_{x' \in \mathbb{R}^n} \{\varphi(x_1,x') + \frac{1}{8}|\varepsilon'-x'|^2\}} e^{\frac{1}{8}|\varepsilon_1 - x_1|^2}\right].
\end{aligned}$$

Using the simple inequality $\inf_{t \in T} \mathbb{E} X_t \geq \mathbb{E} \inf_{t \in T} X_t$ valid for any family of random variables $\{X_t\}_{t \in T}$, we get

$$A \geq \mathbb{E}_{\varepsilon_1, \varepsilon'} e^{\inf_{x',x_1} \left\{\varphi(x_1,x') + \frac{1}{8}|\varepsilon'-x'|^2 + \frac{1}{8}|\varepsilon_1 - x_1|^2\right\}} = \mathbb{E}_{\varepsilon} e^{\inf_{x \in \mathbb{R}^{n+1}} \left\{\varphi(x) + \frac{1}{8}|\varepsilon - x|^2\right\}}.$$

To lower bound B, we first use (3.1) applied to the convex function $x' \mapsto \varphi(x_1, x')$ to see that

$$e^{-\psi(x_1)} = \left(\mathbb{E}_{\varepsilon'} e^{\inf_{x' \in \mathbb{R}^n} \{\varphi(x_1,x') + \frac{1}{8}|\varepsilon'-x'|^2\}}\right)^{-1} \geq \mathbb{E}_{\varepsilon'} e^{-\varphi(x_1,\varepsilon')}.$$

Thus,

$$B = \mathbb{E}_{\varepsilon_1} e^{-\psi(\varepsilon_1)} \geq \mathbb{E}_{\varepsilon_1} \mathbb{E}_{\varepsilon'} e^{-\varphi(\varepsilon_1, \varepsilon')} = \mathbb{E}_{\varepsilon} e^{-\varphi(\varepsilon)}.$$

Putting these two lower bounds into $A \cdot B \leq 1$ shows that (3.1) holds for ψ and hence finishes the proof of the inductive step.

It remains to show (3.1) for $n = 1$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and set

$$\psi(x) = \inf_{y \in \mathbb{R}} \left\{ \varphi(y) + \frac{1}{8}|x - y|^2 \right\}, \qquad x \in \mathbb{R}.$$

Our goal is to show that

$$\frac{e^{\psi(-1)} + e^{\psi(1)}}{2} \cdot \frac{e^{-\varphi(-1)} + e^{-\varphi(1)}}{2} \leq 1.$$

By adding a constant to φ if necessary, without loss of generality we can assume that $\varphi(-1) = 0$. Let $a = \varphi(1)$. By considering $x \mapsto \varphi(-x)$ instead of $x \mapsto \varphi(x)$, we can also assume that $\varphi(-1) \leq \varphi(1)$, that is $a \geq 0$.

Putting $y = -1$ in the definition of ψ gives

$$\psi(-1) \leq \varphi(-1) = 0.$$

Putting in $y = (1 - \lambda) \cdot 1 + \lambda \cdot (-1) = 1 - 2\lambda$ for any $\lambda \in [0, 1]$ gives

$$\begin{aligned} \psi(1) &\leq \varphi((1 - \lambda) \cdot 1 + \lambda \cdot (-1)) + \frac{1}{8}|1 - (1 - 2\lambda)|^2 \\ &\leq (1 - \lambda)\varphi(1) + \lambda\varphi(-1) + \frac{1}{8} \cdot 4\lambda^2 = (1 - \lambda)a + \frac{\lambda^2}{2}. \end{aligned}$$

Taking $\lambda = a$ when $a \leq 1$ and $\lambda = 1$ when $a > 1$ gives the estimate

$$\psi(1) \leq \begin{cases} -\frac{a^2}{2} + a, & a \leq 1, \\ \frac{1}{2}, & a > 1. \end{cases}$$

Therefore, when $a \leq 1$, it suffices to show

$$\frac{1 + e^{-a^2/2 + a}}{2} \cdot \frac{1 + e^{-a}}{2} \leq 1.$$

In fact this holds for any $a \geq 0$: note that the trivial inequality $e^{-a^2/2}(e^a - 1) \leq e^a - 1$ is equivalent to $1 + e^{a - a^2/2} \leq e^a + e^{-a^2/2}$, so

$$\begin{aligned} (1 + e^{a - a^2/2})(1 + e^{-a}) &= 1 + e^{a - a^2/2} + e^{-a}(1 + e^{a - a^2/2}) \\ &\leq 1 + e^{a - a^2/2} + e^{-a}(e^a + e^{-a^2/2}) \\ &= 2 + 2e^{-a^2/2}\cosh a \leq 2 + 2 = 4, \end{aligned}$$

by the well known inequality $\cosh x \leq e^{x^2/2}$, $x \in \mathbb{R}$.

When $a > 1$, it suffices to show

$$\frac{1 + e^{1/2}}{2} \cdot \frac{1 + e^{-a}}{2} \leq 1.$$

The worst case is obviously $a = 1$ and then we have

$$\frac{1+e^{1/2}}{2} \cdot \frac{1+e^{-1}}{2} < \frac{1+1.7}{2} \cdot \frac{1+\frac{1}{2.7}}{2} = \frac{27}{2 \cdot 10} \cdot \frac{37}{2 \cdot 27} = \frac{37}{40} < 1.$$

□

3.4 Log-concave measures

Any log-concave measure satisfies a concentration type inequality for dilations of symmetric convex sets, as established in the following result often called Borell's lemma.

3.12 Theorem (Borell's lemma)**.** *Let μ be a log-concave probability measure on $\mathbb{R}^n$ and let K be a convex symmetric set in $\mathbb{R}^n$ with $\theta = \mu(K) > \frac{1}{2}$. Then for $t \geq 1$,*

$$1 - \mu(tK) \leq \theta \left(\frac{1-\theta}{\theta}\right)^{\frac{t+1}{2}}.$$

Proof. The key observation is that thanks to convexity and symmetry,

$$\frac{t-1}{t+1}K + \frac{2}{t+1}(\mathbb{R}^n \setminus tK) \subset \mathbb{R}^n \setminus K.$$

Indeed, if for some $a \in K$, $\frac{t-1}{t+1}a + \frac{2}{t+1}a' = b$ was in K, then $a' = \frac{t+1}{2t}b + \frac{t-1}{2t}(-a)$ would be in K, too. By log-concavity we thus get

$$\mu(K)^{\frac{t-1}{t+1}}\mu(\mathbb{R}^n \setminus tK)^{\frac{2}{t+1}} \leq \mu(\mathbb{R}^n \setminus K),$$

hence

$$1 - \mu(tK) = \mu(\mathbb{R}^n \setminus tK) \leq \left[\frac{1-\theta}{\theta^{\frac{t-1}{t+1}}}\right]^{\frac{t+1}{2}} = \theta \left(\frac{1-\theta}{\theta}\right)^{\frac{t+1}{2}}.$$

□

3.13 Remark. The right hand side of Borell's lemma is sometimes rewritten as

$$\theta \left(\frac{1-\theta}{\theta}\right)^{\frac{t+1}{2}} = \theta\sqrt{\frac{1-\theta}{\theta}} \left(\frac{1-\theta}{\theta}\right)^{\frac{t}{2}} = \sqrt{\theta(1-\theta)}\sqrt{\frac{1-\theta}{\theta}}^{\,t} \leq \frac{1}{2}\sqrt{1/\theta - 1}^{\,t}$$

and clearly shows the exponential decay since $q = \sqrt{1/\theta - 1} < 1$ for $\theta > 1/2$.

3.14 Remark. Borell's lemma equivalently says that for a log-concave random vector X in $\mathbb{R}^n$ and a semi-norm $\|\cdot\|$ on $\mathbb{R}^n$ we have for $t \geq 1$,

$$\mathbb{P}\left(\|X\| > t\right) \leq \theta \left(\frac{1-\theta}{\theta}\right)^{\frac{t+1}{2}},$$

provided that $\theta = \mathbb{P}\left(\|X\| \leq 1\right) > \frac{1}{2}$.

As an application of this concentration result, we show that the moments of any semi-norm of log-concave vectors do not grow too fast (at most linearly).

3.15 Theorem. *Let X be a log-concave random vector $\mathbb{R}^n$ and let $\|\cdot\|$ be a semi-norm on $\mathbb{R}^n$. Then for $1 \le p < q$,*

$$(\mathbb{E}\|X\|^q)^{1/q} \le 12\frac{q}{p}(\mathbb{E}\|X\|^p)^{1/p}.$$

Proof. By homogeneity, we can assume that $\mathbb{E}\|X\|^p = 1$. Then,

$$\mathbb{P}\left(\|X\| > 4\right) = \mathbb{P}\left(\|X\|^p > 4^p\right) \le 4^{-p}\mathbb{E}\|X\|^p = 4^{-p}.$$

Applying Borell's lemma to $\frac{1}{4}\|\cdot\|$, $\theta = \mathbb{P}\left(\frac{1}{4}\|X\| \le 1\right) \ge 1 - 4^{-p}$, yields

$$\mathbb{P}\left(\frac{1}{4}\|X\| > t\right) \le \sqrt{\theta(1-\theta)}\left(\frac{1}{\theta} - 1\right)^{t/2} \le \left(\frac{1}{1-4^{-p}} - 1\right)^{t/2}, \qquad t > 1.$$

Since, $\frac{1}{1-4^{-p}} - 1 = \frac{4^{-p}}{1-4^{-p}} = \frac{1}{4^p-1} \le \frac{1}{e^p}$,

$$\mathbb{P}\left(\|X\| > 4t\right) \le e^{-pt/2}, \qquad t \ge 1.$$

Thus,

$$\mathbb{E}\|X\|^q = \int_0^\infty qs^{q-1}\mathbb{P}\left(\|X\| > s\right)\mathrm{d}s \le \int_0^4 qs^{q-1}\mathrm{d}s + \int_4^\infty qs^{q-1}\mathbb{P}\left(\|X\| > s\right)\mathrm{d}s.$$

Computing the first integral, changing the variable in the second integral $s = 4t$ and using the above estimate yields

$$\begin{aligned}\mathbb{E}\|X\|^q \le 4^q + \int_1^\infty q4^qt^{q-1}e^{-pt/2}\mathrm{d}t &= 4^q + q4^q\int_1^\infty \left(\frac{2}{p}\right)^q u^{q-1}e^{-u}\mathrm{d}u \\ &\le 4^q\left(1 + \left(\frac{2}{p}\right)^q q\Gamma(q)\right).\end{aligned}$$

Using $q\Gamma(q) \le q^q$ and $(1+x)^{1/q} \le 1 + x^{1/q}$, we finish the proof

$$(\mathbb{E}\|X\|^q)^{1/q} \le 4\left(1 + \left(2\frac{q}{p}\right)^q\right)^{1/q} \le 4\left(1 + 2\frac{q}{p}\right) \le 4\left(\frac{q}{p} + 2\frac{q}{p}\right) = 12\frac{q}{p}.$$

□

3.5 Khinchin–Kahane's inequality

We shall further explore the idea that concentration can be used to obtain moment comparison inequalities and apply this to random signs. Recall classical Bernstein's inequality which says that a weighted sum $\sum_{i=1}^n a_i\varepsilon_i$ of independent random signs ε_i with real coefficients a_i concentrates around its mean (which is 0) with a Gaussian tail,

$$\mathbb{P}\left(|\sum_{i=1}^n a_i\varepsilon_i| > t\right) \le 2e^{-\frac{t^2}{2\sum_{i=1}^n a_i^2}}, \qquad t \ge 0. \tag{3.3}$$

Recall also the following moment comparison which can be shown by a direct computation,

$$\mathbb{E}\left|\sum_{i=1}^n a_i\varepsilon_i\right|^4 \le 3\left(\mathbb{E}\left|\sum_{i=1}^n a_i\varepsilon_i\right|^2\right)^2. \tag{3.4}$$

These two allow us to establish the classical Khinchin inequality. We denote the pth moment of a random variable X by $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$.

3.16 Theorem (Khinchin's inequality). *Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric random signs. For real numbers $a_1, \ldots, a_n$, we have*

$$\left\|\sum_{i=1}^n a_i\varepsilon_i\right\|_p \leq \sqrt{2p}\left\|\sum_{i=1}^n a_i\varepsilon_i\right\|_2, \qquad p \geq 2, \tag{3.5}$$

$$\left\|\sum_{i=1}^n a_i\varepsilon_i\right\|_p \geq \frac{1}{\sqrt{3}}\left\|\sum_{i=1}^n a_i\varepsilon_i\right\|_2, \qquad 1 \leq p \leq 2. \tag{3.6}$$

3.17 Remark. Putting $a_1 = \ldots = a_n = \frac{1}{\sqrt{n}}$ and letting $n \to \infty$, the central limit theorem and the behaviour of Gaussian moments show that the order $\sqrt{p}$ of the constant in (3.5) is best possible.

Proof of Theorem 3.16. By homogeneity, we can assume that $\mathbb{E}|\sum a_i\varepsilon_i|^2 = \sum a_i^2 = 1$. First consider $p \geq 2$. By Bernstein's inequality,

$$\begin{aligned}
\mathbb{E}|\sum a_i\varepsilon_i|^p &= \int_0^\infty pt^{p-1}\mathbb{P}\left(|\sum a_i\varepsilon_i| > t\right)\mathrm{d}t \\
&\leq \int_0^\infty pt^{p-1}2e^{-t^2/2}\mathrm{d}t \\
&= 2^{p/2}p\int_0^\infty u^{p/2-1}e^{-u}\mathrm{d}u \\
&= 2^{p/2}p\Gamma(p/2) \leq 2^{p/2}2(p/2)^{p/2} = 2p^{p/2} \leq (2p)^{p/2}.
\end{aligned}$$

Thus,

$$\|\sum a_i\varepsilon_i\|_p \leq \sqrt{2p} = \sqrt{2p}\|\sum a_i\varepsilon_i\|_2.$$

Now consider $1 \leq p \leq 2$. Let $X = |\sum a_i\varepsilon_i|$. Using Hölder's inequality, we interpolate the $L_2 - L_4$ moment comparison (3.4) to obtain the $L_2 - L_1$ moment comparison,

$$1 = \mathbb{E}X^2 = \mathbb{E}X^{2/3}X^{4/3} \leq (\mathbb{E}X)^{2/3}(\mathbb{E}X^4)^{1/3} \leq 3^{1/3}(\mathbb{E}X)^{2/3}$$

and we finish by crudely bounding the pth moment by the 1st moment,

$$\|\sum a_i\varepsilon_i\|_p \geq \|\sum a_i\varepsilon_i\|_1 = \mathbb{E}X \geq 3^{-1/2} = \frac{1}{\sqrt{3}}\|\sum a_i\varepsilon_i\|_2.$$

□

Talagrand's inequality can be applied to give a simple proof of a substantial generalisation of Khinchin's inequality due to Kahane to vector-valued weights in any normed space.

3.18 Theorem (Kahane's inequality). *Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric random signs. For vectors $v_1, \ldots, v_n$ in a Banach space $(E, \|\cdot\|)$ and $p \geq 1$, we have*

$$\left(\mathbb{E}\|\sum v_i\varepsilon_i\|^p\right)^{1/p} \leq 2\mathbb{E}\|\sum v_i\varepsilon_i\| + 8\sqrt{2}\sqrt{p}\cdot\sigma, \tag{3.7}$$

where $\sigma = \left(\sup\{\sum_{i=1}^n |\varphi(v_i)|^2,\ \varphi \in E^*, \|\varphi\| \le 1\}\right)^{1/2}$. *In particular,*

$$\left(\mathbb{E}\|\sum v_i\varepsilon_i\|^p\right)^{1/p} \le (2 + 8\sqrt{6}\sqrt{p})\mathbb{E}\|\sum v_i\varepsilon_i\|. \tag{3.8}$$

Proof. Define $f(x) = \|\sum_{i=1}^n x_iv_i\|$, $x \in \mathbb{R}^n$. This is clearly a convex function. It is also Lipschitz with constant σ because by duality and the Cauchy-Schwarz inequality,

$$\begin{aligned} f(x) - f(y) \le \|\sum(x_i - y_i)v_i\| &= \sup_{\varphi\in E^*, \|\varphi\|\le 1} |\varphi\left(\sum(x_i - y_i)v_i\right)| \\ &= \sup_{\varphi\in E^*, \|\varphi\|\le 1} |\sum \varphi(v_i)(x_i - y_i)| \\ &\le \sup_{\varphi\in E^*, \|\varphi\|\le 1} \sqrt{\sum|\varphi(v_i)|^2}\sqrt{\sum|x_i - y_i|^2} = \sigma|x - y|. \end{aligned}$$

Let M be the median of $f(\varepsilon)$. By Corollary 3.9, $\mathbb{P}(|f(\varepsilon) - M| > t) \le 4e^{-t^2/(8\sigma^2)}$, so

$$\mathbb{E}|f(\varepsilon) - M|^p \le 4\int_0^\infty pt^{p-1}e^{-t^2/(8\sigma^2)}\mathrm{d}t = 4\sqrt{8\sigma^2}^p\Gamma\left(1 + \frac{p}{2}\right).$$

This, combined with the triangle inequality, gives

$$\left\|\|\sum\varepsilon_iv_i\|\right\|_p - M = \|f(\varepsilon)\|_p - M \le \|f(\varepsilon) - M\|_p \le 4^{1/p}\sqrt{8\sigma^2}\Gamma\left(1 + \frac{p}{2}\right)^{1/p} \le 8\sqrt{2}\sigma\sqrt{p}.$$

It remains to estimate the median. By Chebyshev's inequality, $\frac{1}{2} \le \mathbb{P}(f(\varepsilon) \ge M) \le \frac{1}{M}\mathbb{E}f(\varepsilon)$, so $M \le 2\mathbb{E}f(\varepsilon) = 2\mathbb{E}\|\sum\varepsilon_iv_i\|$, which leads to (3.7).

To obtain (3.8), notice that we can upper bound the parameter σ using the classical Khinchin's inequality (3.6): for any functional $\varphi \in X^*$ with $\|\varphi\| \le 1$, applying (3.6) to $a_i = \varphi(v_i)$ yields

$$\left(\sum|\varphi(v_i)|^2\right)^{1/2} = \left(\mathbb{E}\left|\sum\varphi(v_i)\varepsilon_i\right|^2\right)^{1/2} \le \sqrt{3}\mathbb{E}\left|\sum\varphi(v_i)\varepsilon_i\right| \le \sqrt{3}\mathbb{E}\|\sum\varepsilon_iv_i\|,$$

hence $\sigma \le \sqrt{3}\mathbb{E}\|\sum\varepsilon_iv_i\|$. □

4 Isotropic position

4.1 Isotropic constant

Recall that the **centre of mass (barycentre)** of a subset A of $\mathbb{R}^n$ of positive volume is $\frac{1}{|A|}\int_A x\mathrm{d}x$ and the set A is called **centred** if its centre of mass is at the origin. A convex body K in $\mathbb{R}^n$ is called **isotropic** if

1) $|K| = 1$ and K is centred

2) For every direction $\theta \in S^{n-1}$, $\int_K \langle x, \theta\rangle^2 \mathrm{d}x = L_K^2$ for a positive constant L_K.

The constant L_K is then called the **isotropic constant** of K. For example, every 1-symmetric convex body with volume 1 is isotropic and every orthogonal image of an isotropic convex body is isotropic.

4.1 Remark. Note that under 1), condition 2) is equivalent to each of the following two conditions

2') For every $1 \le i, j \le n$, $\int_K x_i x_j \mathrm{d}x = L_K^2 \delta_{ij}$; in other words, the covariance matrix of the uniform distribution on K equals $L_K^2 \cdot I$

2") For every $n \times n$ real matrix A, $\int_K \langle x, Ax\rangle \mathrm{d}x = L_K^2 \mathrm{tr}(A)$.

Indeed, to see that 2) implies 2'), apply it first to $\theta = e_i$ and then to $\theta = \frac{e_i+e_j}{\sqrt{2}}$. 2') implies 2") by $\langle x, Ax\rangle = \sum A_{ij} x_i x_j$ and 2") implies 2) by applying to the projection $A = \theta\theta^T$.

Intuitively, L_K^2 is the variance of K along any direction. We shall see shortly that one of the central (and unresolved!) questions in asymptotic convex geometry is: "how large can L_K be?"

First we have to establish the existence of the **isotropic position**, meaning that every convex body has an affine (invertible) image which is isotropic. (In general, a **position** of a convex body is any of its affine invertible images). It turns out that such an image is unique up to an orthogonal transformation, so the isotropic position is unique. Of course, there is an affine image of every convex body which has volume one and is centred (satisfies 1)).

4.2 Theorem (Existence)**.** *Let K be a centred convex body in $\mathbb{R}^n$. Then there is an invertible linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that TK is isotropic.*

Proof. We can assume that $|K| = 1$. Take $A = [\int_K x_i x_j \mathrm{d}x]_{i,j=1}^n$. This is a symmetric positive definite matrix because

$$\langle Au, u\rangle = \sum_{i,j} A_{ij} u_i u_j = \int_K \sum_{i,j} x_i x_j u_i u_j \mathrm{d}x = \int_K \langle x, u\rangle^2 \mathrm{d}x,$$

which is nonnegative and nonzero (K has nonempty interior). By the spectral theorem, $A = U^T DU$ for an orthogonal matrix U and a positive diagonal matrix D. Take $T = D^{-1/2}U$ to get that

$$\int_{TK} x_i x_j \mathrm{d}x = |\det T| \int_K (Ty)_i (Ty)_j \mathrm{d}y = |\det T| \int_K \sum_{k,l} T_{ik} y_k T_{jl} y_l \mathrm{d}y$$
$$= |\det T| [\underbrace{TAT^T}_{I}]_{i,j} = |\det T| \delta_{ij},$$

so TK is isotropic. □

4.3 Theorem (Uniqueness)**.** *Let K be a centred convex body of volume 1 in $\mathbb{R}^n$. Let*

$$m_K = \inf_{T \in SL(n)} \int_{TK} |x|^2 \mathrm{d}x.$$

Then for $T \in SL(n)$, a position $K' = TK$ of K is isotropic if and only if $\int_{K'} |x|^2 \mathrm{d}x = m_K$. Moreover, if K_1, K_2 are two such positions, then $K_2 = UK_1$ for some $U \in O(n)$.

Proof. Let K' be an isotropic position of K. Then for any $T \in SL(n)$,

$$\int_{TK'} |x|^2 \mathrm{d}x = \int_{K'} |Ty|^2 \mathrm{d}y = \int_{K'} \langle y, T^T Ty \rangle \mathrm{d}y = L_K^2 \mathrm{tr}(T^T T)$$
$$\geq L_K^2 n (\det(T^T T))^{1/n}$$
$$= L_K^2 \cdot n = \int_{K'} |x|^2 \mathrm{d}x,$$

so $\int_{K'} |x|^2 \mathrm{d}x$ achieves the infimum of the quantity $\int_{TK'} |x|^2 \mathrm{d}x$ over all $T \in SL(n)$.

If some K_1 and K_2 are isotropic positions of K, then $K_2 = TK_1$ and in the above AM-GM inequality for T, $\mathrm{tr}(T^T T) \geq (\det(T^T T))^{1/n}$ we have equality, so $T^T T = I$, hence $T \in O(n)$. □

4.4 Remark. By the above characterision of the isotropic position as minimising the second moment of the Euclidean norm, the isotropic position and hence the isotropic constant depend only on the affine class of a convex body: two convex bodies have the same isotropic position up to an orthogonal transformation, if and only if one is an invertible affine image of the other. Moreover, for any convex body K,

$$L_K^2 = \frac{1}{n} \min_{T \in GL(n)} \left\{ \frac{1}{|T\tilde{K}|^{1+2/n}} \int_{T\tilde{K}} |x|^2 \mathrm{d}x \right\}, \tag{4.1}$$

where $\tilde{K}$ is the translate of K with barycentre at the origin, $\tilde{K} = K - \mathrm{bar}(K) = K - \frac{1}{|K|} \int_K x \mathrm{d}x$. Explanation: let positive λ be such that $|\lambda \tilde{K}| = 1$; by Theorem 4.3, any $T_0 \in SL(n)$ which minimises $\int_{T_0(\lambda)\tilde{K}} |x|^2 \mathrm{d}x$ is such that $T_0(\lambda \tilde{K})$ is isotropic, so

$$L_K^2 = L_{T_0(\lambda\tilde{K})}^2 = \frac{1}{n} \int_{T_0(\lambda\tilde{K})} |x|^2 \mathrm{d}x = \frac{1}{n} \min_{T \in SL(n)} \left\{ \int_{T(\lambda\tilde{K})} |x|^2 \mathrm{d}x \right\}$$
$$= \frac{1}{n} \min_{T \in SL(n)} \left\{ \lambda^{n+2} \int_{T(\tilde{K})} |x|^2 \mathrm{d}x \right\}.$$

Since $\lambda = 1/|\tilde{K}|^{1/n} = 1/|T\tilde{K}|^{1/n}$ for any $T \in SL(n)$, the last expression can be written as the right hand side of (4.1).

It is easy to give a sharp lower bound for the isotropic constant by pushing mass as near the origin as possible.

4.5 Theorem. *For a convex body K in $\mathbb{R}^n$, we have $L_K \geq L_{B_2^n}$.*

Proof. We can assume that K is isotropic. Let r_n be such that $B = r_n B_2^n$ be isotropic. Note that since both K and B have the same volume, $K \setminus B$ and $B \setminus K$ have the same volume. We have,

$$\begin{aligned} nL_K^2 = \int_K |x|^2 \mathrm{d}x &= \int_{K\cap B} |x|^2 + \int_{K\setminus B} |x|^2 \\ &\geq \int_{K\cap B} |x|^2 + \int_{K\setminus B} r_n^2 \\ &= \int_{K\cap B} |x|^2 + \int_{B\setminus K} r_n^2 \\ &\geq \int_{K\cap B} |x|^2 + \int_{B\setminus K} |x|^2 = \int_B |x|^2 = nL_{B_2^n}^2. \end{aligned}$$

□

4.6 Remark. The isotropic constant of the Euclidean ball can be found using integration in polar coordinates,

$$L_{B_2^n}^2 = \frac{1}{n}\int_{r_n B_2^n} |x|^2 \mathrm{d}x = \frac{1}{n}\int_0^{r_n} r^2 r^{n-1} |S^{n-1}| \mathrm{d}r = \frac{1}{n+2} r_n^{n+2} |B_2^n|.$$

By the choice of r_n, $1 = |r_n B_2^n| = r_n^n |B_2^n|$ and recall $|B_2^n| = \frac{\sqrt{\pi}^n}{\Gamma(1+\frac{n}{2})}$, so we obtain

$$L_{B_2^n}^2 = \frac{1}{n+2}|B_2^n|^{-2/n} = \frac{1}{n+2}\frac{\Gamma\left(1+\frac{n}{2}\right)^{2/n}}{\pi} \xrightarrow[n\to\infty]{} \frac{1}{2e\pi} \approx 0.058...$$

The fundamental conjecture in convex geometry called the **slicing or hyperplane conjecture** says that the isotropic constant is bounded above by an absolute constant.

4.7 Conjecture (Slicing problem)**.** There is a universal constant $C > 0$ such that for every $n \geq 1$ and every convex body K in $\mathbb{R}^n$, the isotropic constant of K is bounded by C, $L_K \leq C$.

There is no known example of a convex body K with $L_K > 1$. It is suspected that among symmetric convex bodies, the worst case (the one with largest isotropic constant) would be the cube. Currently, the best known bound is due to Klartag from 2006 which says that $L_K \leq Cn^{1/4}$ for every convex body K in $\mathbb{R}^n$, where C is a universal constant. The slicing conjecture is known to hold for certain classes of convex bodies, for example for the unconditional convex bodies. We shall see very soon (in the next chapter) a proof of the weaker bound $L_K \leq Cn^{1/2}$.

4.2 Why “slicing”?

At first glance, it is not evident what Conjecture 4.7 or the isotropic constant has to do with slicing a convex body. We shall see a reason by showing that Conjecture 4.7 is equivalent to the following one.

4.8 Conjecture (Slicing/Hyperplane conjecture)**.** There is a universal constant $c > 0$ such that for every $n \geq 1$ and every *centred* convex body K in $\mathbb{R}^n$ with $|K| = 1$, there is a direction $\theta \in S^{n-1}$ such that $|K \cap \theta^\perp| \geq c$ (there is a central hyperplane slice of constant volume).

Given a centred convex body K of volume 1 and a direction $\theta \in S^{n-1}$, consider the function of the $n-1$-dimensional volume of the slice of K with a hyperplane perpendicular to θ passing through $t\theta$, $t \in \mathbb{R}$,

$$f(t) = |K \cap (t\theta + \theta^\perp)|.$$

By Brunn's principle, this function is a $\frac{1}{n-1}$-concave on its support. In particular, f is log-concave. Moreover,

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = |K| = 1,$$
$$\int_{-\infty}^{\infty} tf(t)\mathrm{d}t = \int_K \langle x, \theta\rangle \mathrm{d}x = 0,$$
$$\int_{-\infty}^{\infty} t^2 f(t)\mathrm{d}t = \int_K \langle x, \theta\rangle^2 \mathrm{d}x$$
$$f(0) = |K \cap \theta^\perp|.$$

Recall that for a centred log-concave function, by Corollary 2.24, the three quantities $\int f$, $\int t^2 f(t)$ and $f(0)$ are related and we have

$$\frac{1}{12e^2}\frac{1}{f(0)^2} \leq \int t^2 f(t)\mathrm{d}t \leq 2\frac{1}{f(0)^2}.$$

This proves the following result.

4.9 Theorem. *For a centred convex body K in $\mathbb{R}^n$ with volume 1 and $\theta \in S^{n-1}$, we have*

$$\frac{1}{2\sqrt{3}e}\frac{1}{|K \cap \theta^\perp|} \leq \left(\int_K \langle x, \theta\rangle^2 \mathrm{d}x\right)^{1/2} \leq \sqrt{2}\frac{1}{|K \cap \theta^\perp|}.$$

In particular, if K is isotropic, the volumes of all central sections are essentially the same, within absolute constants of $1/L_K$,

$$\frac{1}{2\sqrt{3}e}\frac{1}{L_K} \leq |K \cap \theta^\perp| \leq \sqrt{2}\frac{1}{L_K}.$$

This immediately shows that the existence of a section of large volume gives an upper bound for the isotropic constant.

Conjecture 4.8 implies Conjecture 4.7. Suppose K is isotropic. By Conjecture 4.8, there is θ_0 such that $|K \cap \theta_0^\perp| \geq c$ with some absolute constant $c > 0$, so $L_K \leq \frac{\sqrt{2}}{c}$. □

To see the other implication, we need a lemma saying that centred convex bodies have directions of variance not exceeding the isotropic constant squared.

4.10 Lemma. *Let K be a centred convex body in $\mathbb{R}^n$ with volume 1. Then there is $\theta \in S^{n-1}$ such that*

$$\int_K \langle x, \theta\rangle^2 \mathrm{d}x \leq L_K^2.$$

Proof. Let $M_K = [\int_K x_i x_j \mathrm{d}x]_{i,j\leq n}$ be the covariance matrix of K and consider the ellipsoid of inertia of K,

$$\mathcal{E}_K = \{x \in \mathbb{R}^n,\ \langle x, M_K x\rangle \leq 1\}.$$

We shall compute its volume in two ways. Recall (1.2), which gives

$$|\mathcal{E}_K| = (\det M_K)^{-1/2}|B_2^n|.$$

Moreover, for any $T \in SL(n)$, $M_{TK} = TM_KT^T$, so $\det M_K = \det M_{TK}$. Choosing T such that TK is isotropic, we have $M_{TK} = L_K^2 I$ and therefore $\det M_K = L_K^{2n}$, which gives

$$|\mathcal{E}_K| = L_K^{-n}|B_2^n|.$$

On the other hand, integrating in polar coordinates,

$$|\mathcal{E}_K| = \int_{S^{n-1}} \int_0^{1/\|\theta\|_{\mathcal{E}_K}} t^{n-1}|S^{n-1}|\mathrm{d}\sigma(\theta)\mathrm{d}t = |B_2^n| \int_{S^{n-1}} \|\theta\|_{\mathcal{E}_K}^{-n}\mathrm{d}\sigma(\theta).$$

Putting the two together yields

$$L_K^{-n} = \int_{S^{n-1}} \|\theta\|_{\mathcal{E}_K}^{-n}\mathrm{d}\sigma(\theta).$$

In particular, there is $\theta_0 \in S^{n-1}$ such that $\|\theta_0\|_{\mathcal{E}_K}^{-n} \geq L_K^{-n}$. Recalling that the norm on $\mathcal{E}_K$ is given by M_K, see (1.3), we get

$$\int_K \langle x, \theta_0\rangle^2 \mathrm{d}x = \langle \theta_0, M_K\theta_0\rangle = \|\theta_0\|_{\mathcal{E}_K}^2 \leq L_K^2.$$

□

Conjecture 4.7 implies Conjecture 4.8. Suppose K is a centred convex body with volume 1. Let $\theta_0 \in S^{n-1}$ be a direction guaranteed by Lemma 4.10. Then by Theorem 4.9,

$$\frac{1}{2\sqrt{3}e}\frac{1}{|K \cap \theta_0^\perp|} \leq \left(\int_K \langle x, \theta_0\rangle^2 \mathrm{d}x\right)^{1/2} \leq L_K,$$

therefore, if $L_K \leq C$ for an absolute constant C, then $|K \cap \theta_0^\perp| \geq \frac{1}{2\sqrt{3}eC}$. □

4.3 Inradius and outerradius

We show that the inradius and outerradius of an isotropic convex body is related to its isotropic constant and deduce an easy upper bound for the isotropic constant. The inradius is the largest possible radius of a ball contained in the body and the outerradius is the smallest possible radius of a ball containing the body.

4.11 Theorem. *Let K be an isotropic convex body in $\mathbb{R}^n$. Then,*

$$K \subset 2\sqrt{3}e \cdot nL_K B_2^n.$$

Moreover, if K is symmetric, then

$$K \supset L_K B_2^n.$$

Proof. For the upper bound, fix $\theta \in S^{n-1}$ and choose x_0 such that $h_K(\theta) = \langle x_0, \theta\rangle$. Then the cone $C_\theta = \operatorname{conv}\{K \cap \theta^\perp, x_0\}$ is contained in K, thus looking at volumes gives

$$1 = |K| \geq |C_\theta| = \frac{1}{n} h_K(\theta)|K \cap \theta^\perp|.$$

This and Corollary 4.9 yield

$$h_K(\theta) \leq n\frac{1}{|K \cap \theta^\perp|} \leq 2\sqrt{3}enL_K = h_{2\sqrt{3}enL_K B_2^n}(\theta).$$

Since θ is arbitrary, we get $K \subset 2\sqrt{3}e \cdot nL_K B_2^n$.

For the lower bound, by symmetry $|\langle x, \theta\rangle| \leq h_K(\theta)$ for any $\theta \in S^{n-1}$ and $x \in K$, therefore for any $\theta \in S^{n-1}$,

$$L_K = \left(\int_K \langle x, \theta\rangle \mathrm{d}x\right)^{1/2} \leq h_K(\theta),$$

which gives $L_K B_2^n \subset K$. $\square$

4.12 Corollary. *Let K be a symmetric isotropic convex body in $\mathbb{R}^n$. Then,*

$$L_K \leq |B_2^n|^{-1/n} \leq \frac{1}{2}\sqrt{n}.$$

Proof. Since $K \supset L_K B_2^n$, looking at volumes gives $1 = |K| \geq L_K^n |B_2^n|$. Recall the formula for the volume of the Euclidean ball (1.1) to get

$$|B_2^n|^{-1/n} = \frac{1}{\sqrt{\pi}}\Gamma\left(1 + \frac{n}{2}\right)^{1/n}.$$

When $n \geq 2$, by $\Gamma(1+x) \leq x^x$, $x \geq 1$, this is upper bounded by $\frac{1}{\sqrt{\pi}}\left(\frac{n}{2}\right)^{\frac{n}{2}\frac{1}{n}} = \frac{1}{\sqrt{2\pi}}\sqrt{n} < \frac{1}{2}\sqrt{n}$. When $n = 1$, $|B_2^n|^{-1/n} = \frac{1}{2}$. $\square$

4.13 Remark. The symmetry assumption in Theorem 4.11 and hence in Corollary 4.12 can be omitted.

4.4 Isotropic log-concave measures

A random vector X in $\mathbb{R}^n$ is **isotropic** if

1) X is centred, that is $\mathbb{E}X = 0$

2) $\mathrm{Cov}(X) = \mathbb{E}XX^T = I$, equivalently, $\mathbb{E}\langle X, \theta\rangle^2 = 1$ for every $\theta \in S^{n-1}$.

For instance, a standard Gaussian vector $X \sim N(0, I)$ is isotropic.

How should we define the isotropic constant of isotropic random vectors? Consider the example of a random vector X uniformly distributed on a convex body K in $\mathbb{R}^n$. Let $f(x) = \frac{1}{|K|}\mathbf{1}_K(x)$ be the density of X. We have

$$\mathbb{E}X = \int_{\mathbb{R}^n} xf(x)\mathrm{d}x = \frac{1}{|K|}\int_K x\mathrm{d}x$$

and

$$\mathbb{E}\langle x, \theta\rangle^2 = \int_{\mathbb{R}^n} \langle x, \theta\rangle^2 f(x)\mathrm{d}x = \frac{1}{|K|}\int_K \langle x, \theta\rangle^2 \mathrm{d}x.$$

Suppose X is isotropic. Then 1) implies that K is centred and 2) gives $\int_K \langle x, \theta\rangle^2 \mathrm{d}x = |K|$, for every $\theta \in S^{n-1}$. Therefore, if we take $\lambda > 0$ such that $|\lambda K| = 1$, that is $\lambda = |K|^{-1/n}$, then λK is an isotropic convex body and

$$L_K^2 = \int_{\lambda K} \langle x, \theta\rangle^2 \mathrm{d}x = \lambda^{n+2}\int_K \langle x, \theta\rangle^2 \mathrm{d}x = \lambda^{n+2}|K| = |K|^{-2/n}.$$

In other words, $L_K = \frac{1}{|K|^{1/n}} = \sup_{x\in\mathbb{R}^n} f(x)^{1/n}$. This motives our definition.

The **isotropic constant** L_X of a log-concave isotropic random vector X in $\mathbb{R}^n$ with density f_X is defined as

$$L_X = \|f_X\|_\infty^{1/n}.$$

In particular, if X is symmetric, $L_X = f_X(0)^{1/n}$.

Suppose now that we have a log-concave random vector X with density f_X, which is not necessarily isotropic. We would still like to have an expression for its isotropic constant. Let $b = \mathbb{E}X$ so that $X - b$ is centred and take a matrix A such that $Y = A(X - b)$ is isotropic, that is $I = \mathrm{Cov}(A(X - b)) = A\mathrm{Cov}(X - b)A^T = A\mathrm{Cov}(X)A^T$ (why does such A exist?). Then $L_X = L_Y = \|f_Y\|_\infty^{1/n}$. We have

$$f_Y(x) = \det(A^{-1})f_{X-b}(A^{-1}x) = (\det A)^{-1}f_X(A^{-1}(x + b))$$

and $1 = \det(A\mathrm{Cov}(X)A^T) = (\det A)^2 \det(\mathrm{Cov}(X))$, so

$$L_X = \|f_Y\|_\infty^{1/n} = (\det A)^{-1/n}\|f_X\|_\infty^{1/n} = (\det \mathrm{Cov}(X))^{\frac{1}{2n}}\|f_X\|_\infty^{1/n}.$$

4.14 Corollary. *Let X be a log-concave random vector X in $\mathbb{R}^n$ with density f. Then the isotropic constant of X equals*

$$L_X = (\det \mathrm{Cov}(X))^{\frac{1}{2n}}\|f\|_\infty^{1/n}. \tag{4.2}$$

Since log-concave distributions naturally generalise uniform distributions on convex sets, it is reasonable to ask in the spirit of the slicing problem whether the isotropic constants of the former are also uniformly bounded.

4.15 Conjecture (Slicing problem'). There is a positive constant C such that for every n and every continuous log-concave random vector X in $\mathbb{R}^n$, $L_X \leq C$.

The slicing problem for convex bodies (Conjecture 4.7) and this presumably stronger conjecture are in fact equivalent. There is a construction which produces a convex body given a log-concave vector with the isotropic constants of the two different by at most a constant factor. It relies on Ball's inequality (Theorem 2.25).

5 John's position

5.1 Maximal volume ellipsoids

Existence of extremal objects is often times useful. In this chapter, we consider ellipsoids of maximal volume contained in convex bodies. It turns out this has interesting and important applications.

We start with existence and uniqueness of such ellipsoids.

5.1 Lemma. *Given a convex body K in $\mathbb{R}^n$, there exists a unique ellipsoid of maximal volume inscribed in K.*

Proof. To show the existence, consider the set $\mathcal{A}$ of all ellipsoids contained in K

$$\mathcal{A} = \{(b,T),\ b \in \mathbb{R}^n, T \text{ is an } n \times n \text{ positive definite real matrix },b + TB_2^N \subset K\}.$$

This is a bounded set (if $b+TB_2^n \subset K \subset RB_2^n$, then $b \in RB_2^n$, $TB_2^n \subset RB_2^n$, so $\|T\|_{\mathrm{op}} \leq R$) which is closed, hence compact. Therefore, the supremum of $|b+TB_2^n| = (\det T)|B_2^n|$ is attained on $\mathcal{A}$ (because det is continuous), which show that there is an ellipsoid of maximal volume in K.

To address the uniqueness, suppose there are two ellipsoids $\mathcal{E}_1$ and $\mathcal{E}_2$ in K of maximal volume. Without loss of generality, say $\mathcal{E}_1 = B_2^n$ and $\mathcal{E}_2 = b + TB_2^n$. Since $|\mathcal{E}_1| = |\mathcal{E}_2|$, we have $\det T = 1$. Since $\mathcal{E}_1, \mathcal{E}_2 \subset K$, by convexity,

$$\mathcal{E} = \frac{b}{2} + \frac{I+T}{2}B_2^n = \frac{\mathcal{E}_1 + \mathcal{E}_2}{2} \subset K,$$

so $\mathcal{E}$ is another ellipsoid in K and looking at volumes, by the maximality of $\mathcal{E}_1$ and $\mathcal{E}_2$, $\det\left(\frac{I+T}{2}\right) \leq 1$. On the other hand, if we denote the eigenvalues of T by t_i, then by the AM-GM inequality,

$$\det\left(\frac{I+T}{2}\right)^{1/n} = \left[\prod \frac{1+t_i}{2}\right]^{1/n} \geq \left[\prod \frac{1}{2}\right]^{1/n} + \left[\prod \frac{t_i}{2}\right]^{1/n} = \frac{1}{2} + \frac{1}{2}(\det T)^{1/n} = 1,$$

thus we have equality here, which is the case if and only if the t_i are constant, so they are all 1, that is $T = I$, or $\mathcal{E}_2 = b + B_2^n$. If $\mathcal{E}_1 \neq \mathcal{E}_2$, then $b \neq 0$, but then we can dilate the ellipsoid $\frac{b}{2} + B_2^n \subset \mathrm{conv}\{B_2^n, b + B_2^n\} \subset K$ a bit along the direction b to get an ellipsoid in K of a larger volume. □

5.2 Corollary. *Given a convex body K in $\mathbb{R}^n$, there exists a unique ellipsoid of minimal volume containing K.*

Proof. Use duality (apply Lemma 5.1 to polars). □

We say that a convex body is in **John's position** if B_2^n is its ellipsoid of maximal volume. Since B_2^n is an invertible affine image of any ellipsoid, by Lemma 5.1 any convex body can be put (by an invertible affine map) in John's position.

John showed that when a body is in John's position, there are points which make this fact much more workable. Ball showed that the sort of opposite statement also holds, that is in order to check whether B_2^n is the maximal volume ellipsoid, it suffices to exhibit contact points. A point x is a **contact point** of a body K and B_2^n, if $x \in \partial B_2^n \cap \partial K$, or in other words, $|x| = 1 = \|x\|_K$. We shall assume symmetry in both of these results.

5.3 Theorem (John). *If B_2^n is the maximal volume ellipsoid inside a symmetric convex body K in $\mathbb{R}^n$, then there exist contact points $x_1, \ldots, x_m$ of K and B_2^n and positive numbers $c_1, \ldots, c_m$ with $m \leq \binom{n+1}{2} + 1$ such that for every $x \in \mathbb{R}^n$,*

$$x = \sum_{i=1}^{m} c_i \langle x, x_i \rangle x_i. \tag{5.1}$$

5.4 Remark. If $B_2^n \subset K$ and x is a contact point of K and B_2^n, then x is also a contact point of the polar, $\|x\|_{K^\circ} = 1$. Indeed, since $B_2^n \subset K$, if H is a supporting hyperplane at x of K, then H is also a supporting hyperplane at x of B_2^n, but there is only one choice for the latter, namely $x + x^\perp$, so $H = x + x^\perp$. Therefore,

$$\|x\|_{K^\circ} = h_K(x) = \sup_{y \in K} \langle x, y \rangle \leq \sup_{y \in H^-} \langle x, y \rangle \leq \langle x, x \rangle = 1.$$

On the other hand, we trivially have $\|x\|_{K^\circ} = h_K(x) = \sup_{y \in K} \langle x, y \rangle \geq \langle x, x \rangle = 1$ by picking $y = x$.

5.5 Remark. Condition (5.1) can be equivalently stated in terms of matrices as

$$I = \sum_{i=1}^{m} c_i x_i x_i^T. \tag{5.2}$$

Taking trace gives in particular that

$$\sum_{i=1}^{m} c_i = n. \tag{5.3}$$

Moreover,

$$|x|^2 = \langle x, x \rangle = \left\langle x, \sum c_i \langle x, x_i \rangle x_i \right\rangle = \sum_{i=1}^{m} c_i \langle x, x_i \rangle^2. \tag{5.4}$$

Proof of Theorem 5.3. Looking at (5.2) and (5.3), we see that in fact we want to show that

$$\frac{1}{n} I \in \text{conv}\{xx^T,\ x \text{ is a contact point}\}$$

where xx^T can be treated as elements of $\mathbb{R}^{\binom{n+1}{2}}$ (because the positive semi-definite matrices can be viewed as a subset of $\mathbb{R}^{\binom{n+1}{2}}$). Then, by Carathéodory's theorem 1.1, we know that it is enough to take $m \leq \binom{n+1}{2} + 1$ contact points.

If $\frac{1}{n} I$ is not in the convex hull of the contact points, it can be separated from it, meaning there is an $n \times n$ symmetric matrix ϕ such that

$$\left\langle \phi, \frac{1}{n} I \right\rangle < \alpha < \langle \phi, xx^T \rangle, \qquad x \in \partial B_2^n \cap \partial K.$$

By considering $\phi - \frac{1}{n}(\mathrm{tr}\phi)I$, we can assume that $\langle \phi, \frac{1}{n}I \rangle = 0$ and $\mathrm{tr}\phi = 0$, so

$$\langle \phi x, x \rangle = \sum_{i,j} \phi_{i,j} x_i x_j = \langle \phi, xx^T \rangle > \alpha$$

for all contact points for some $\alpha > 0$.

For $\delta > 0$ small enough (so small that $I + \delta\phi$ is positive definite), consider the (nondegenerate) ellipsoid

$$\mathcal{E}_\delta = \{x \in \mathbb{R}^n, \ \langle x, (I + \delta\phi)x \rangle \le 1\}.$$

Note that

$$|\mathcal{E}_\delta| = (\det(I + \delta\phi))^{-1/2} |B_2^n| > \left(\frac{1}{n} \mathrm{tr}(I + \delta\phi) \right)^{-n/2} |B_2^n| = |B_2^n|$$

by the AM-GM inequality (which is strict because otherwise all the eigenvalues of ϕ are the same and zero as $\mathrm{tr}\phi = 0$, but ϕ is nonzero). This means that $\mathcal{E}_\delta$ is an ellipsoid of a larger volume that B_2^n. We reach a desired contradiction, if we show that $\mathcal{E}_\delta$ is in K for δ small enough.

To this end, we show that for every unit vector v we have $\frac{v}{\|v\|_K} \notin \mathcal{E}_\delta$. Let

$$U = \{u \in S^{n-1}, \ u \text{ is a contact point of } K \text{ and } B_2^n\}$$

be the set of all contact points. First consider the unit vectors which are away from the contact points

$$V = \left\{ x \in S^{n-1}, \ \mathrm{dist}(x, U) \ge \frac{\alpha}{2\|\phi\|} \right\}.$$

This is a compact set. Let $d = \max\{\|x\|_K, x \in V\}$. Note that $d < 1$ because $B_2^n \subset K$. Let $\lambda = \min_{x \in V} \langle \phi x, x \rangle$. Since $\mathrm{tr}\phi = 0$ and ϕ is nonzero, it has at least one negative and one positive eigenvalue. In particular, there is a vector w such that $\langle \phi w, w \rangle < 0$. Then for every $u \in U$,

$$0 > \langle \phi w, w \rangle = \langle \phi u, u \rangle + \langle \phi u, w - u \rangle + \langle \phi(w - u), u \rangle > \alpha - 2|w - u| \|\phi\|,$$

hence $|w - u| > \frac{\alpha}{2\|\phi\|}$ implying $w \in V$ and $\lambda < 0$. Take $\delta < \frac{1-d^2}{|\lambda|}$. Then for $v \in V$,

$$\left\| \frac{v}{\|v\|_K} \right\|_{\mathcal{E}_\delta}^2 = \left\langle (I + \delta\phi) \frac{v}{\|v\|_K}, \frac{v}{\|v\|_K} \right\rangle = \frac{1 + \delta \langle \phi v, v \rangle}{\|v\|_K^2} \ge \frac{1 + \delta\lambda}{\|v\|_K^2} \ge \frac{1 + \delta\lambda}{d^2} > 1,$$

that is $\frac{v}{\|v\|_K} \notin \mathcal{E}_\delta$.

Now suppose $v \in S^{n-1} \setminus V$. There is a contact point $u \in U$ such that $|u - v| < \frac{\alpha}{2\|\phi\|}$ and we have

$$\begin{aligned} |\langle (I + \delta\phi)v, v \rangle - \langle (I + \delta\phi)u, u \rangle| &= \delta \, |\langle \phi v, v \rangle - \langle \phi u, u \rangle| \\ &\le \delta \, |\langle \phi v, v \rangle - \langle \phi v, u \rangle| + \delta \, |\langle \phi v, u \rangle - \langle \phi u, u \rangle| \\ &\le 2\delta \|\phi\| |u - v| < \delta\alpha. \end{aligned}$$

Since $v \in B_2^n \subset K$, $\|v\|_K \leq 1$ and we get

$$\left\| \frac{v}{\|v\|_K} \right\|_{\mathcal{E}_\delta}^2 \geq \|v\|_{\mathcal{E}_\delta}^2 = \langle (I+\delta\phi)v, v\rangle \geq \langle (I+\delta\phi)v, v\rangle - \delta\alpha > 1 + \delta\alpha - \delta\alpha = 1,$$

thus $\frac{v}{\|v\|_K} \notin \mathcal{E}_\delta$ in this case as well. ☐

5.6 Theorem (Ball). *Let K be a symmetric convex body in $\mathbb{R}^n$ such that it contains B_2^n and there are some contact points $u_1, \ldots, u_m \in \partial B_2^n \cap \partial K$ and weights $c_1, \ldots, c_m > 0$ such that*

$$I = \sum_{i=1}^m c_i u_i u_i^T.$$

Then B_2^n is the maximal volume ellipsoid in K.

Proof. Instead of K, consider the polyhedron given by the hyperplanes tangent to the unit ball at contact points,

$$L = \{y \in \mathbb{R}^n,\ \forall\, i \leq m\ \langle u_i, y\rangle \leq 1\}.$$

Particularly, the u_i are also contact points of L and B_2^n. If $y \in K$, then $\langle u_i, y\rangle \leq \|u_i\|_{K^\circ}\|y\|_K = 1$, so $K \subset L$ and it is enough to show that B_2^n is the maximal volume ellipsoid in L. Take an ellipsoid

$$\mathcal{E} = \left\{ x \in \mathbb{R}^n,\ \sum_{i=1}^n \frac{\langle x, v_i\rangle^2}{\alpha_i^2} \right\}$$

with $(v_i)_{i=1}^n$ being an orthonormal basis and $\alpha_i > 0$. Suppose $\mathcal{E} \subset L$. Let $y_i = \sum_j \alpha_j \langle u_i, v_j\rangle v_j$. Since

$$\sum_k \frac{\langle y_i, v_k\rangle^2}{\alpha_k^2} = \sum_k \langle u_i, v_k\rangle^2 = |u_i|^2 = 1,$$

we get $y_i \in \mathcal{E} \subset L$ and thus $\langle y_i, u_i\rangle \leq \|y_i\|_L \|u_i\|_{L^\circ} \leq 1$. Therefore,

$$\begin{aligned}\sum_j \alpha_j = \sum_j \alpha_j |v_j|^2 = \sum_{i,j} c_i \alpha_j \langle u_i, v_j\rangle^2 &= \sum_i c_i \left\langle \sum_j \alpha_j \langle u_i, v_j\rangle v_j, u_i \right\rangle \\ &= \sum_i c_i \langle y_i, u_i\rangle \leq \sum_i c_i = n.\end{aligned}$$

By the AM-GM inequality,

$$|\mathcal{E}| = \left(\prod \alpha_j\right) |B_2^n| \leq |B_2^n|,$$

so B_2^n is the maximal volume ellipsoid. ☐

5.7 Remark. Theorems 5.3 and 5.6 give a characterisation that for a symmetric convex body K which contains B_2^n, B_2^n is the maximal volume ellipsoid in K if and only if for some unit vectors $u_1, \ldots, u_m \in \partial K$ and positive weights $c_1, \ldots, c_m$ we have $I = \sum_{i=1}^m c_i u_i u_i^T$. If K is not necessarily symmetric, the same remains true after adding the condition that $\sum_{i=1}^m c_i u_i = 0$.

5.2 Applications

Banach-Mazur distance

Our first application draws from the fact that symmetric convex bodies in John's position have a not too large outerradius (cf. Theorem 4.11).

5.8 Theorem. *For a symmetric convex body K in $\mathbb{R}^n$ which is in John's position we have*

$$B_2^n \subset K \subset \sqrt{n} B_2^n.$$

5.9 Corollary. *The Banach-Mazur distance of any symmetric convex body in $\mathbb{R}^n$ to the unit ball B_2^n is at most $\sqrt{n}$.*

Proof. Consider $T \in GL(n)$ such that TK is in John's position. Then $B_2^n \subset TK \subset \sqrt{n}B_2^n$, so recalling the definition of the Banach-Mazur distance, we see that indeed $d_{BM}(K, B_2^n) \leq \sqrt{n}$. $\square$

Proof of Theorem 5.8. Note that by (5.4) and (5.3) for any $x \in K$, we have

$$|x|^2 = \sum_i c_i \langle x, u_i \rangle^2 \leq \sum_i c_i (\|x\|_K \|u_i\|_{K^\circ})^2 \leq \sum_i c_i = n,$$

where the u_i are contact points and c_i are weights from John's theorem 5.3. $\square$

Thanks to the (multiplicative) "triangle inequality" for the Banach-Mazur distance, we also get that for any two symmetric convex bodies K and L in $\mathbb{R}^n$,

$$d_{BM}(K, L) \leq n.$$

This bound is sharp in a sense that for n large enough, there are symmetric convex bodies K and L in $\mathbb{R}^n$ such that $d_{BM}(K, L) > cn$, for an absolute constant $c > 0$ (Gluskin's theorem).

Circumscribed cube

Our second application is about circumscribing a small cube around a symmetric convex body in John's position. It turns out to be possible to fit a certain $\sqrt{n}$-dimensional slice of an n-dimensional body into a cube of constant side length. This result will be crucial in the next chapter when we discuss almost Euclidean sections of convex bodies.

5.10 Theorem (Dvoretzky-Rogers factorisation)**.** *Suppose that B_2^n is the maximal volume ellipsoid of a symmetric convex body K in $\mathbb{R}^n$. There are $s = \left\lfloor \frac{\sqrt{n}}{4} \right\rfloor$ orthogonal unit vectors $z_1, \ldots, z_s$ such that for every reals $a_1, \ldots, a_s$, we have*

$$\frac{1}{2} \max_{i \leq s} |a_i| \leq \left\| \sum_{i=1}^{s} a_i z_i \right\|_K \leq \sqrt{\sum_{i=1}^{s} a_i^2}.$$

5.11 Remark. Equivalently, the assertion says that there is a subspace H of dimension $s = \left\lfloor \frac{\sqrt{n}}{4} \right\rfloor$ and an orthogonal map U such that

$$B_2^n \cap H \subset K \cap H \subset 2(UB_\infty^n \cap H).$$

Proof. Let $x_1, \dots, x_m$ be contact points and $c_1, \dots, c_m$ positive weights guaranteed by John's theorem 5.3 such that

$$I = \sum_{i=1}^{m} c_i x_i x_i^T.$$

Let $z_1 = x_1$. We select the remaining vectors z_i in the following greedy procedure: suppose that orthogonal unit vectors $z_1, \dots, z_k$ have been selected and consider the projection P onto $(\mathrm{span}\{z_1, \dots, z_k\})^\perp$. Let $l \le m$ be an index such that

$$|Px_l| = \max_{j \le m} |Px_j|.$$

Note that then

$$n - k = \mathrm{tr}P = \mathrm{tr}(PI) = \sum c_i \mathrm{tr}(Px_i x_i^T) = \sum c_i \mathrm{tr}\langle x_i, Px_i\rangle = \sum c_i |Px_i|^2 \\ \le |Px_l|^2 n.$$

Some explanation: since P is a projection, $P^T = P$ and $P^2 = P$, so $P = P^T P$ and

$$\langle x_i, Px_i\rangle = \langle x_i, P^T P x_i\rangle = \langle Px_i, Px_i\rangle$$

In the last inequality we used the choice of l and (5.3). Rearranging we get

$$|Px_l|^2 \ge \frac{n-k}{n}.$$

In particular, Px_l is nonzero. Set

$$z_{k+1} = \frac{Px_l}{|Px_l|}.$$

Clearly, $z_1, \dots, z_k, z_{k+1}$ are unit orthogonal vectors. We shall show that in this inductive greedy procedure we also have

$$\|z_k\|_{K^\circ} \le 2, \qquad k \le \frac{\sqrt{n}}{4}.$$

Since z_1 was chosen among contact points, $\|z_1\|_{K^\circ} = 1$. Suppose that $\|z_j\|_{K^\circ} \le 2$ for $j \le k$. Let us write x_l selected in the $k+1$ step as

$$x_l = Px_l + (I - P)x_l = |Px_l| z_{k+1} + \sum_{j=1}^{k} \alpha_j z_j,$$

for some $\alpha_j \in \mathbb{R}$. Using orthogonality,

$$1 = |x_l|^2 = |Px_l|^2 + \sum_{j=1}^{k} \alpha_j^2,$$

hence

$$\sum_{j=1}^{k} \alpha_j^2 = 1 - |Px_l|^2 \le \frac{k}{n}.$$

Therefore,

$$\|z_{k+1}\|_{K^\circ} = \left\| \frac{x_l - \sum_j \alpha_j z_j}{|Px_l|} \right\|_{K^\circ} \le \sqrt{\frac{n}{n-k}} \left(\|x_l\|_{K^\circ} + \sum_j |\alpha_j| \|z_j\|_{K^\circ} \right).$$

Since x_l is a contact point, $\|x_l\|_{K^\circ} = 1$ and by the inductive assumption $\|z_j\|_{K^\circ} \le 2$. By the Cauchy-Schwarz inequality, $\sum_j |\alpha_j| \le \sqrt{k}\sqrt{\sum_j \alpha_j^2} \le \frac{k}{\sqrt{n}}$. Putting it all together yields

$$\|z_{k+1}\|_{K^\circ} \le \sqrt{\frac{n}{n-k}} \left(1 + \frac{2k}{\sqrt{n}} \right).$$

The right hand side is clearly an increasing function of k. Plugging in $k = \frac{\sqrt{n}}{4}$ it becomes $\sqrt{\frac{1}{1-\frac{1}{4\sqrt{n}}}} \cdot \frac{3}{2} \le \sqrt{\frac{1}{1-\frac{1}{4\sqrt{1}}}} \cdot \frac{3}{2} = \sqrt{3} < 2$.

We have constructed at least $s = \left\lfloor \frac{\sqrt{n}}{4} \right\rfloor$ orthogonal unit vectors $z_1, \ldots, z_s$ such that $\|z_j\|_{K^\circ} \le 2$. Consider $z = \sum_{j=1}^s a_j z_j$, $a_1, \ldots, a_s \in \mathbb{R}$. Observe that for any $j \le s$,

$$|a_j| = |\langle z, z_j \rangle| \le \|z\|_K \|z_j\|_{K^\circ} \le 2\|z\|_K,$$

which gives the left inequality of the assertion. The right one is clear because of orthogonality and $B_2^n \subset K$,

$$\|z\|_K \le \|z\|_{B_2^n} = |z| = \sqrt{\sum_{j=1}^{s} a_j^2}.$$

□

Reverse isoperimetry

Recall that the classical isoperimetry (see Theorem 2.6) says that among all sets with fixed volume, spheres have the smallest surface area. Let us consider the reverse problem: among all sets with fixed volume, which ones have the largest surface area? A quick thought reveals that pancakes can in fact have arbitrarily large surface area having their volume fixed. What if we ask the same question modulo affine invariance, meaning we consider sets the same when they are invertible affine images of one another?

5.12 Theorem (Ball). *(i) Let K be a convex body in $\mathbb{R}^n$ and let Δ be a regular n-dimensional simplex. Then there is an affine image $\tilde{K}$ of K such that*

$$|\tilde{K}| = |\Delta| \qquad \textit{and} \qquad |\partial \tilde{K}| \le |\partial \Delta|.$$

(ii) If K is in addition symmetric, then there is a linear image $\tilde{K}$ of K such that

$$|\tilde{K}| = |B_\infty^n| \qquad \textit{and} \qquad |\partial \tilde{K}| \le |\partial B_\infty^n|.$$

The **volume ratio** of a convex body K in $\mathbb{R}^n$ is defined as

$$\mathrm{vr}(K) = \left(\frac{|K|}{|\mathcal{E}|}\right)^{1/n}, \qquad \mathcal{E} \text{ is the maximal volume ellipsoid in } K.$$

Note that this is an affine invariant quantity. The reverse isoperimetry theorem due to Ball follows from his theorem concerning sets having maximal volume ratio.

5.13 Theorem (Ball)**.** *(i) Among all convex bodies in $\mathbb{R}^n$, the n-dimensional regular simplex has the largest volume ratio.*
(ii) Among all symmetric convex bodies in $\mathbb{R}^n$, the cube B_∞^n has the largest volume ratio.

We shall only consider the symmetric case (the nonsymmetric case requires further, a bit technical, considerations).

Proof of Theorem 5.12 (ii) from Theorem 5.13 (ii). Let $\tilde{K}$ be the affine image of K such that for some positive λ, λB_2^n is the maximal volume ellipsoid in $\tilde{K}$ and $|\tilde{K}| = |B_\infty^n|$. Since $B_2^n \subset \frac{1}{\lambda}\tilde{K}$, we have

$$|\partial\tilde{K}| = \liminf_{\varepsilon\to 0+} \frac{|\tilde{K} + \varepsilon B_2^n| - |\tilde{K}|}{\varepsilon} \leq \liminf_{\varepsilon\to 0+} \frac{|\tilde{K} + \frac{\varepsilon}{\lambda}\tilde{K}| - |\tilde{K}|}{\varepsilon} = |\tilde{K}|\frac{n}{\lambda} = |B_\infty^n|\frac{n}{\lambda}.$$

Note that $n|B_\infty^n| = 2n \cdot 2^{n-1} = |\partial B_\infty^n|$. By Theorem 5.13,

$$\frac{1}{\lambda^n}\frac{|\tilde{K}|}{|B_2^n|} = \frac{|\tilde{K}|}{|\lambda B_2^n|} = \mathrm{vr}(K)^n \leq \mathrm{vr}(B_\infty^n)^n = \frac{|B_\infty^n|}{|B_2^n|},$$

so canceling $|\tilde{K}| = |B_\infty^n|$ and $|B_2^n|$ gives $\frac{1}{\lambda} \leq 1$, thus $|\partial\tilde{K}| \leq |\partial B_\infty^n|$. $\square$

The proof of Theorem 5.13 about maximising the volume ratio goes from Ball's geometric form of the Brascamp-Lieb inequality, which we leave for now and show it later, together with the reversal due to Barthe.

5.14 Theorem (Ball/Brascamp-Lieb)**.** *If some unit vectors $u_1, \ldots, u_m$ in $\mathbb{R}^n$ and positive numbers $c_1, \ldots, c_m$ satisfy*

$$I = \sum_{i=1}^m c_i u_i u_i^T,$$

then for any integrable functions $f_1, \ldots, f_m : \mathbb{R} \to [0, \infty)$, we have

$$\int_{\mathbb{R}^n} \prod_{i=1}^m (f_i(\langle x, u_i\rangle))^{c_i} \mathrm{d}x \leq \prod_{i=1}^m \left(\int_{\mathbb{R}} f_i\right)^{c_i}.$$

Proof of Theorem 5.13 (ii) from Theorem 5.14. Since the maximal volume ellipsoid of B_∞^n is B_2^n and the volume ration is affine invariant, we need to show that if B_2^n is the maximal volume ellipsoid in K, then $|K| \leq |B_\infty^n|$. In that case, by John's theorem 5.3, there are contact points $u_1, \ldots, u_m$ and positive numbers $c_1, \ldots, c_m$ such that $I =$

$\sum c_i x_i x_i^T$. By symmetry, $K \subset \bigcap_{i\leq m}\{x \in \mathbb{R}^n,\ |\langle x, u_i\rangle| \leq 1\}$, thus from Theorem 5.14 and (5.3),

$$\begin{aligned}|K| = \int_{\mathbb{R}^n} \mathbf{1}_K(x)\mathrm{d}x \leq \int_{\mathbb{R}^n} \prod_{i\leq m} \mathbf{1}_{[-1,1]}(\langle x, u_i\rangle)^{c_i}\mathrm{d}x &\leq \prod_{i\leq m}\left(\int \mathbf{1}_{[-1,1]}\right)^{c_i}\\ &= \prod_{i\leq m} 2^{c_i} = 2^n = |B_\infty^n|.\end{aligned}$$

□

6 Almost Euclidean sections

Recall the Dvoretzky-Rogers theorem 5.10 which says that every symmetric convex body K in $\mathbb{R}^n$ admits a subspace H of dimension roughly $\sqrt{n}$ such that $B_2^H \subset K \cap H \subset 2B_\infty^H$, where B_2^H is the unit Euclidean ball in H and B_∞^H is the cube in H with respect to a certain orthonormal basis. Grothendieck asked whether B_∞^H can be replaced with B_2^H so that we have a sort of matching lower and upper bound, possibly lowering the dimension of H but still going to infinity as $n \to \infty$. Dvoretzky answered this question positively. The optimal dimension dependence was established by Milman using concentration of measure. The result itself as well as its influential proof are cornerstones of asymptotic convex geometry.

In this chapter we only focus on what is true when the dimension is large enough and do not care about values of absolute constants. For convenience, $c, C, c_1, C_1, \dots$ always denote positive universal constants, values of which may change from one occurrence to another.

6.1 Dvoretzky's theorem

The goal of this section is to prove the following quantitative version of Dvoretzky's theorem.

6.1 Theorem. *There is an absolute constant c such that for every $\varepsilon \in (0,1)$, every symmetric convex body K in $\mathbb{R}^n$ has a $(1+\varepsilon)$-Euclidean section of dimension $k \geq \frac{c\varepsilon^2}{\log\frac{1}{c\varepsilon}}\log n$, that is there is a k-dimensional subspace F of $\mathbb{R}^n$ and a constant $C > 0$ such that*

$$\frac{1}{1+\varepsilon}C(B_2^n \cap F) \ \subset\ K \cap F \ \subset\ (1+\varepsilon)C(B_2^n \cap F).$$

This is a stronger statement than

$$d_{BM}(K \cap F, B_2^n \cap F) \leq (1+\varepsilon)^2$$

which amounts to saying that there is an invertible linear map $T \in GL(n)$ such that

$$T(B_2^n \cap F) \subset K \cap F \subset (1+\varepsilon)^2 T(B_2^n \cap F),$$

that is K has a $(1+\varepsilon)$-ellipsoidal section ($T(B_2^n \cap F)$ is an ellipsoid).

On the other hand, ellipsoids admit exact Euclidean sections of a proportional dimension, as shown in the next lemma.

6.2 Lemma. *Let $\mathcal{E}$ be a centred ellipsoid in $\mathbb{R}^n$. There is a $\lceil\frac{n}{2}\rceil$-dimensional subspace F of $\mathbb{R}^n$ such that $\mathcal{E} \cap F$ is a Euclidean ball.*

Proof. By possibly rotating, we can assume that

$$\mathcal{E} = \{x \in \mathbb{R}^n,\ \sum_{i=1}^n \alpha_i x_i^2 \leq 1\}$$

with $0 < \alpha_1 \le \alpha_2 \le \ldots \le \alpha_n$. Take $c = \mathrm{Med}((\alpha_i)_{i=1}^n)$ and F to be the subspace of the solutions to the system of equations

$$\{\forall\ i \le \lfloor \frac{n}{2} \rfloor \ \ \sqrt{c - \alpha_i} x_i = \sqrt{\alpha_{n-i} - c} x_{n-i}$$

Then on F, $\alpha_i x_i^2 + \alpha_{n-i} x_{n-i}^2 = c(x_i^2 + x_{n-i}^2)$ for all $i \le \lfloor \frac{n}{2} \rfloor$, so on F, $\sum_i \alpha_i x_i^2 = \sum_i c x_i^2$ (note that when n is odd, $c = \alpha_{(n+1)/2}$). This means that $\mathcal{E} \cap F$ is a ball. □

6.3 Remark. Lemma 6.2 means that if we find a $(1+\varepsilon)$-ellipsoidal section, we also get a $(1+\varepsilon)$-Euclidean section of a dimension possibly smaller, but at most by a factor of 2. Therefore, to prove Theorem 6.1, where we do not care about absolute constants, it suffices to get a suitable ellipsoidal section.

We set off to prove Dvoretzky's theorem. We fix some notation. For a normed space $X = (\mathbb{R}^n, \|\cdot\|)$ and its unit ball K we introduce two parameters: the mean of the norm, **mean norm**,

$$M = M_X = M_K = \int_{S^{n-1}} \|\theta\| \mathrm{d}\sigma(\theta)$$

and its Lipschitz constant

$$b = \inf\{t > 0,\ \forall\ x \in \mathbb{R}^n\ \|x\| \le t|x|\} = \sup_{x \in S^{n-1}} \|x\|$$

that is the smallest constant b such that $B_2^n \subset bK$. Throughout this chapter, we shall write $\|\cdot\|$ for $\|\cdot\|_K$ (unless it is ambiguous).

The quantity $\frac{M}{b}$ plays a crucial role in obtaining large dimensional Euclidean sections, as explained in the next theorem due to Milman.

6.4 Theorem (Milman)**.** *If K is a symmetric convex body in $\mathbb{R}^n$, then for every $\varepsilon \in (0,1)$ and $k \le \frac{c\varepsilon^2}{\log \frac{1}{c\varepsilon}} n \left(\frac{M}{b}\right)^2$, there is a subset Γ of the set $G_{n,k}$ of k-dimensional subspaces of $\mathbb{R}^n$ of Haar measure $\nu_{n,k}(\Gamma) \ge 1 - \exp\left\{-c\varepsilon^2 n \left(\frac{M}{b}\right)^2\right\}$ such that*

$$\forall F \in \Gamma \quad \frac{1}{1+\varepsilon} \frac{1}{M} (B_2^n \cap F) \ \subset\ K \cap F \ \subset\ (1+\varepsilon) \frac{1}{M} (B_2^n \cap F).$$

Here c is an absolute positive constant.

We will easily deduce Dvoretzky's theorem provided that we know $n\left(\frac{M}{b}\right)^2$ is at least roughly $\log n$. This is the case for bodies in John's position as clarified in the following lemma.

6.5 Lemma. *If B_2^n is the maximal volume ellipsoid in a convex body K in $\mathbb{R}^n$, then*

$$\frac{M_K}{b} \ge c\sqrt{\frac{\log n}{n}},$$

where $c > 0$ is an absolute constant.

Proof of Dvoretzky's theorem 6.1 from Milman's theorem 6.4. For a symmetric convex body K in $\mathbb{R}^n$, take a linear map $T \in GL(n)$ such that TK is in John's position. By Milman's theorem and Lemma 6.5 applied to TK, we get a $(1+\varepsilon)$-Euclidean section of TK of dimension

$$k_0 \geq \left\lfloor \frac{c\varepsilon^2}{\log\frac{1}{c\varepsilon}} n \left(\frac{M}{b}\right)^2 \right\rfloor \geq \left\lfloor \frac{c\varepsilon^2}{\log\frac{1}{c\varepsilon}} \log n \right\rfloor.$$

This gives a $(1+\varepsilon)$-ellipsoidal section of K of dimension k_0, which by Lemma 6.2 (see also Remark 6.3) gives a $(1+\varepsilon)$-Euclidean section of K of dimension $\lceil \frac{k_0}{2} \rceil$. □

Proof of Lemma 6.5. We have $B_2^n \subset K$, in other words, $\|x\| \leq |x|$ for every $x \in \mathbb{R}^n$, so $b \leq 1$. It suffices to show that M is large. Let $u_1, \ldots, u_k$, $k = \lfloor \frac{\sqrt{n}}{4} \rfloor$ be orthogonal unit vectors from the Dvoretzky-Rogers factorisation 5.10 applied to K, that is $\|\sum_{i \leq k} a_i u_i\| \geq \frac{1}{2}\max_{i\leq k}|a_i|$. Extend $(u_i)_{i\leq k}$ to an orthonormal basis $(u_i)_{i\leq n}$ of $\mathbb{R}^n$. Then by rotational invariance,

$$\begin{aligned} M = \int_{S^{n-1}} \|\theta\| \mathrm{d}\sigma(\theta) &= \int_{S^{n-1}} \|\sum \theta_i e_i\| \mathrm{d}\sigma(\theta_1, \ldots, \theta_n) \\ &= \int_{S^{n-1}} \|\sum \theta_i \varepsilon_i u_i\| \mathrm{d}\sigma(\theta_1, \ldots, \theta_n). \end{aligned}$$

for any choice of signs $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$. In particular, averaging over a random vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ of independent random signs yields

$$M = \int_{S^{n-1}} \mathbb{E}_\varepsilon \|\sum \varepsilon_i \theta_i u_i\| \mathrm{d}\sigma(\theta_1, \ldots, \theta_n).$$

By independence and the triangle inequality,

$$\mathbb{E}_\varepsilon \|\sum \varepsilon_i \theta_i u_i\| \geq \mathbb{E}_{(\varepsilon_i)_{i\leq k}} \|\sum_{i\leq k} \varepsilon_i \theta_i u_i + \mathbb{E}_{(\varepsilon_i)_{i>k}} \sum_{i>k} \theta_i \varepsilon_i u_i\| = \mathbb{E}_\varepsilon \|\sum_{i\leq k} \varepsilon_i \theta_i u_i\|,$$

thus

$$M \geq \int_{S^{n-1}} \mathbb{E}_\varepsilon \|\sum_{i\leq k} \varepsilon_i \theta_i u_i\| \mathrm{d}\sigma(\theta) = \int_{S^{n-1}} \|\sum_{i\leq k} \theta_i u_i\| \mathrm{d}\sigma(\theta) \geq \frac{1}{2} \int_{S^{n-1}} \max_{i\leq k} |\theta_i| \mathrm{d}\sigma(\theta).$$

Since $k \geq c\sqrt{n}$, it remains to show the following lemma. □

6.6 Lemma. *For every $k \leq n$,*

$$\int_{S^{n-1}} \max_{i\leq k} |\theta_i| \mathrm{d}\sigma(\theta) \geq c\sqrt{\frac{\log k}{n}},$$

where $c > 0$ is a universal constant.

Proof. Let X be a random vector uniformly distributed on S^{n-1}. It is easier to work with a standard Gaussian vector G in $\mathbb{R}^n$ rather than X because we can use independence and the two are related (recall Theorem A.2): $X \sim \frac{G}{|G|}$. We have

$$\int_{S^{n-1}} \max_{i\leq k} |\theta_i| \mathrm{d}\sigma(\theta) = \mathbb{E}\max_{i\leq k}|X_i| = \mathbb{E}\frac{1}{|G|}\max_{i\leq k}|G_i|.$$

By Chebyshev's inequality,

$$\mathbb{P}\left(|G| \geq \sqrt{3n}\right) \leq \frac{1}{3n}\mathbb{E}|G|^2 = \frac{1}{3}.$$

By independence,

$$\mathbb{P}\left(\max_{i\leq k}|G_i| \leq c\sqrt{\log k}\right) = \prod_{i\leq k}\mathbb{P}\left(|G_i| \leq c\sqrt{\log k}\right) = \mathbb{P}\left(|G_1| \leq c\sqrt{\log k}\right)^k.$$

By a direct computation,

$$\begin{aligned}\mathbb{P}\left(|G_1| \leq c\sqrt{\log k}\right) = 1 - \frac{2}{\sqrt{2\pi}}\int_{c\sqrt{\log k}}^{\infty} e^{-t^2/2}\mathrm{d}t &\leq 1 - \frac{2}{\sqrt{2\pi}}\int_{c\sqrt{\log k}}^{c\sqrt{\log k}+1} e^{-t^2/2}\mathrm{d}t \\ &\leq 1 - \frac{2}{\sqrt{2\pi}}e^{-(c\sqrt{\log k}+1)^2/2}\mathrm{d}t.\end{aligned}$$

For $k \geq 2$ and, say $c = \frac{\sqrt{2}}{10}$, we get $c\sqrt{\log k} + 1 \leq \sqrt{2\log k}$ and

$$\mathbb{P}\left(|G_1| \leq c\sqrt{\log k}\right)^k \leq \left(1 - \sqrt{\frac{\pi}{2}}\frac{1}{k}\right)^k < e^{-\sqrt{\frac{\pi}{2}}} < \frac{1}{2}.$$

Putting these together, the union bound gives

$$\mathbb{P}\left(|G| < \sqrt{3n},\ \max_{i\leq k}|G_i| > c\sqrt{\log k}\right) \geq 1 - \frac{1}{3} - \frac{1}{2} = \frac{1}{6},$$

consequently,

$$\mathbb{E}\frac{1}{|G|}\max_{i\leq k}|G_i| \geq \frac{1}{6}\frac{1}{\sqrt{3n}}c\sqrt{\log k}.$$

□

6.7 Remark. Regardless the position, we always have $\frac{M}{b} \geq \frac{c}{\sqrt{n}}$ with a universal constant $c > 0$. This is essentially because, by the definition of b, $B_2^n \subset bK$ and there is a contact point $u \in \partial bK \cap S^{n-1}$, so bK is contained is the symmetric slab $\{x \in \mathbb{R}^n,\ \langle x, u\rangle \leq 1\}$. It remains to compare the norms.

Our last task in this section is to prove Milman's theorem. We fix $\varepsilon \in (0,1)$ and a symmetric convex body K in $\mathbb{R}^n$. We write in short $\|\cdot\|_K = \|\cdot\|$. We want to find a subset Γ of subspaces (of large dimension) of large measure for which the sections of K are $(1+\varepsilon)$-Euclidean, that is

$$\forall F \in \Gamma \quad \frac{1}{1+\varepsilon}\frac{1}{M}(B_2^n \cap F) \ \subset\ K \cap F \ \subset\ (1+\varepsilon)\frac{1}{M}(B_2^n \cap F).$$

or

$$\forall F \in \Gamma \quad \frac{1}{1+\varepsilon}M \leq \|x\| \leq (1+\varepsilon)M, \qquad x \in S_F = S^{n-1} \cap F.$$

The argument is based on concentration of measure on the sphere (Corollary 3.2) and approximation by nets (Lemma B.3). Recall the crucial parameters: the mean norm, $M = \int_{S^{n-1}}\|\theta\|\mathrm{d}\sigma(\theta)$ and the Lipschitz constant, smallest b such that $\|x\| \leq b|x|$ for all $x \in \mathbb{R}^n$.

Proof of Milman's theorem 6.4

Step 1 (Majority of rotations send unit vectors close to $M\partial K$). For unit vectors $y_1, \ldots, y_m$ with $m \le c_1 e^{c_1 \varepsilon^2 n}$, there is a set B of orthogonal maps, $B \subset O(n)$ of Haar measure $\nu(B) \ge 1 - e^{-c_2 \varepsilon^2 n}$ such that

$$\forall U \in B \ \forall j \le m \quad M - b\varepsilon \le \|Uy_j\| \le M + b\varepsilon. \tag{6.1}$$

Explanation: consider the set

$$A = \{x \in S^{n-1}, \ M - b\varepsilon \le \|x\| \le M + b\varepsilon\}$$

and apply the concentration for the 1-Lipschitz function $\frac{1}{b}\|x\|$ on S^{n-1} around its mean $\frac{M}{b}$ (Corollary 3.2) to get

$$\sigma(A) \ge 1 - Ce^{-c\varepsilon^2 n}.$$

For every $j \le m$ take the set of "good" orthogonal maps for y_j,

$$B_j = \{U \in O(n), \ M - b\varepsilon \le \|Uy_j\| \le M + b\varepsilon\}$$

and let $B = \cap_{j \le m} B_j$. Of course, (6.1) holds for this set B. Since $\nu(B_j) = \sigma(A)$, the union bound gives

$$\nu(B^c) \le \sum_{j \le m} \nu(B_j^c) \le m \cdot Ce^{-c\varepsilon^2 n} \le c_1 C e^{(c_1 - c)\varepsilon^2 n} \le e^{-c_2 \varepsilon^2 n}.$$

Step 2 (Random subspaces work well for nets). If $\left(1 + \frac{2}{\delta}\right)^k \le c_1 e^{c_1 \varepsilon^2 n}$, then there is a set $\Gamma \subset G_{n,k}$ of k-dimensional subspaces of Haar measure $\nu(\Gamma) \ge 1 - e^{-c_2 \varepsilon^2 n}$ such that for every $F \in \Gamma$ there is a δ-net N_F of S_F (for the Euclidean distance) with

$$M - b\varepsilon \le \|x\| \le M + b\varepsilon, \qquad x \in N_F. \tag{6.2}$$

Explanation: let $F_0 = \text{span}\{e_1, \ldots, e_k\}$ and take a δ-net $N_0 = \{y_1, \ldots, y_m\}$ of S_{F_0} with $m \le \left(1 + \frac{2}{\delta}\right)^k$ (Lemma B.3). Take a set $B \subset O(n)$ provided by Step 1 for the vectors $y_1, \ldots, y_m$ and for every $U \in B$ define $F_U = UF_0$, $N_F = UN_0$ (note that $S_{F_U} = US_{F_0}$. Clearly, the choice $\Gamma = \{F_U, \ U \in B\} \subset G_{n,k}$ is as desired and $\nu(\Gamma) = \nu(B)$.

Step 3 (From nets to whole spheres). For a set Γ from Step 2, for every $F \in \Gamma$,

$$\frac{1 - 2\delta}{1 - \delta} M - \frac{b\varepsilon}{1 - \delta} \le \|x\| \le \frac{M + b\varepsilon}{1 - \delta}, \qquad x \in S_F. \tag{6.3}$$

Explanation: for the upper bound, we want to show that $A = \sup_{y \in S_F} \|y\| \le \frac{M + b\varepsilon}{1 - \delta}$. Consider $x \in S_F$ along with its approximation $x_0 \in N_F$ from a δ-net N_F of S_F such that $|x - x_0| \le \delta$. From Step 2, $\|x_0\| \le M + b\varepsilon$, so

$$\|x\| \le \|x - x_0\| + \|x_0\| \le \left\| \frac{x - x_0}{|x - x_0|} \right\| |x - x_0| + M + b\varepsilon \le A\delta + M + b\varepsilon.$$

Taking the supremum over $x \in S_F$, we get $A \le A\delta + M + b\varepsilon$ as needed.

For the lower bound, a similar argument gives

$$\begin{aligned} \|x\| \ge \|x_0\| - \|x - x_0\| &\ge M - b\varepsilon - \left\| \frac{x - x_0}{|x - x_0|} \right\| |x - x_0| \\ &\ge M - b\varepsilon - A\delta \\ &\ge M - b\varepsilon - \frac{M + b\varepsilon}{1 - \delta}\delta = \frac{1 - 2\delta}{1 - \delta} M - \frac{b\varepsilon}{1 - \delta}. \end{aligned}$$

Step 4 (Adjusting parameters). Given $\varepsilon_0 \in (0, 1)$, we use Step 2 and 3 with

$$\delta = \frac{\varepsilon_0}{6} \qquad \text{and} \qquad \varepsilon = \frac{M}{b}\delta = \frac{M}{b}\frac{\varepsilon_0}{6}.$$

We need to guarantee that $\left(1 + \frac{2}{\delta}\right)^k \le c_1 e^{c_1 \varepsilon^2 n}$, that is

$$\left(1 + \frac{12}{\varepsilon_0}\right)^k \le c_1 e^{\frac{c_1}{36}\varepsilon_0^2 n \left(\frac{M}{b}\right)^2},$$

which holds as long as

$$k \le \frac{c\varepsilon_0^2}{\log \frac{1}{c\varepsilon_0}} n \left(\frac{M}{b}\right)^2.$$

We get a set $\Gamma \subset G_{n,k}$ of subspaces of Haar measure

$$\nu(\Gamma) \ge 1 - \exp\{-c_2 \varepsilon^2 n\} = 1 - \exp\left\{-c\varepsilon_0^2 n \left(\frac{M}{b}\right)^2\right\}.$$

such that for every subspace $F \in \Gamma$, we have (6.3), that is

$$\frac{1 - 2\delta}{1 - \delta} M - \frac{b\varepsilon}{1 - \delta} \le \|x\| \le \frac{M + b\varepsilon}{1 - \delta}, \qquad x \in S_F.$$

We check that

$$\frac{1 - 2\delta}{1 - \delta} M - \frac{b\varepsilon}{1 - \delta} \ge \frac{1}{1 + \varepsilon_0} M$$

and

$$\frac{M + b\varepsilon}{1 - \delta} \le (1 + \varepsilon_0) M.$$

This finishes the proof of Milman's theorem 6.4. □

6.2 Critical dimension

For an n-dimensional normed space $X = (\mathbb{R}^n, \|\cdot\|)$ we define its **critical dimension** $k(X)$ as the largest integer $k_0 \le n$ for which

$$\nu \left\{ F \in G_{n,k},\ \frac{1}{2} M|x| \le \|x\| \le 2M|x| \ \forall x \in F \right\} \ge 1 - \frac{k_0}{n + k_0}, \qquad k = 1, \dots, k_0.$$

In words, this is the largest dimension of 2-Euclidean sections existing for prevailing subspaces. We also set $\tilde{k}(X)$ to be the largest integer $k_0 \le n$ for which

$$\nu \left\{ F \in G_{n,k},\ \frac{1}{2} M|x| \le \|x\| \le 2M|x| \ \forall x \in F \right\} \ge \frac{1}{2}, \qquad k = 1, \dots, k_0.$$

Note that $\tilde{k}(X) \ge k(X)$.

6.8 Remark. By Milman's theorem 6.4,

$$\tilde{k}(X) \geq cn \left(\frac{M}{b}\right)^2.$$

Indeed, if $n\left(\frac{M}{b}\right)^2 \leq 1/c$, there is nothing to prove. Otherwise, apply Theorem 6.4 to, say $\varepsilon = \frac{1}{2}$ to get that there is an integer $k_0 \geq c_1 n\left(\frac{M}{b}\right)^2$ such that for every $k \leq k_0$ there is a set Γ of k-dimensional subspaces such that $\nu(\Gamma) \geq 1 - e^{-c_2 n\left(\frac{M}{b}\right)^2} \geq 1 - e^{-c_2/c} \geq 1/2$ and for every $F \in \Gamma$ and $x \in F$,

$$\frac{2}{3}M|x| \leq \|x\| \leq \frac{3}{2}M|x|.$$

Thus $\tilde{k}(X) \geq k_0$.

If a multiple of the unit ball of X is in John's position, then

$$k(X) \geq cn \left(\frac{M}{b}\right)^2.$$

Indeed, apply Theorem 6.4 as above to $\varepsilon = \frac{1}{2}$ to get that for $k_0 = \lfloor c_1 n\left(\frac{M}{b}\right)^2 \rfloor$ and every $k \leq k_0$, there is a set Γ of k-dimensional subspaces with $3/2$-Euclidean sections such

$$\nu(\Gamma) \geq 1 - e^{-c_2 n\left(\frac{M}{b}\right)^2}.$$

For $C = \frac{c_2}{c_1}$ we get $c_2 n\left(\frac{M}{b}\right)^2 \geq Ck_0$, so

$$\nu(\Gamma) \geq 1 - e^{-Ck_0} \geq 1 - \frac{k_0}{e^{Ck_0} + k_0}.$$

By Lemma 6.5, $n\left(\frac{M}{b}\right)^2 \geq c\log n$, so (possibly increasing C), $e^{Ck_0} \geq n$ and consequently,

$$\nu(\Gamma) \geq 1 - \frac{k_0}{n + k_0}.$$

Thus $k(X) \geq k_0$.

The mysterious threshold $\frac{k}{n+k}$ in the definition of the critical dimension is partially explained by the following theorem, a strong reversal of the previous remark.

6.9 Theorem (Milman-Schechtman)**.** *There is a universal constant $C > 0$ such that for every n dimensional normed space X, its critical dimension satisfies*

$$k(X) \leq 8n \left(\frac{M}{b}\right)^2.$$

Proof. Let k be equal to $k(X)$, so that we can write $n = kt + r$ for integers $t \geq 0$ and $k > r \geq 0$. Take orthogonal subspaces $E_1, \ldots, E_t$ of dimension k and an orthogonal subspace E_{t+1} of dimension r such that $\mathbb{R}^n = \sum_{i=1}^{t+1} E_i$ (if $r = 0$, we only need to take $E_1, \ldots, E_t$). By the definition of the critical dimension, for each i

$$\nu\{U \in O(n),\ UE_i \text{ gives a 2-Euclidean section}\} \geq 1 - \frac{k}{n+k}.$$

Note that, if $r > 0$, $t = \frac{n-r}{k} < \frac{n}{k}$, $(t+1)\frac{k}{n+k} < 1$, and if $r = 0$, $t\frac{k}{n+k} < 1$, therefore, by the union bound, there is $U \in O(n)$ such that for each i, UE_i gives a 2-Euclidean section, that is

$$\forall i\ \forall x \in UE_i \qquad \frac{1}{2}M|x| \le \|x\| \le 2M|x|.$$

For every $x \in \mathbb{R}^n$, we write $x = \sum_{i=1}^{t+1} x_i$ with $x_i \in E_i$ so that $|x|^2 = \sum |x_i|^2$ and by the Cauchy-Schwarz inequality we obtain

$$\|x\| \le \sum \|x_i\| \le 2M|x_i| \le 2M\sqrt{t+1}|x|.$$

This implies $b \le 2M\sqrt{t+1}$, thus

$$n\left(\frac{M}{b}\right)^2 \ge n\frac{1}{4(t+1)} > \frac{1}{4}n\frac{k}{n+k} \ge \frac{1}{8}k.$$

□

Application to polytopes

Note that the cube B_∞^n has 2^n vertices and $2n$ facets, the cross-polytope B_1^n has $2n$ vertices and 2^n facets. It turns out that symmetric polytopes have etiher a lot of facets or vertices, which is not the case without symmetry because in $\mathbb{R}^n$, for instance an n-simplex has $n+1$ vertices and $n+1$ facets ($n-1$-dimensional faces).

6.10 Theorem. *If P is a symmetric polytope in $\mathbb{R}^n$ with $f(P)$ facets and $v(P)$ vertices, then*

$$\log f(P) \ \cdot \ \log v(P) \ge cn$$

with a universal constant $c > 0$.

We shall obtain this from the following theorem and lemma.

6.11 Theorem. *For every n-dimensional normed space X whose unit ball is in John's position,*

$$k(X)k(X^*) \ge cn,$$

where $c > 0$ is a universal constant and X^* is the dual.

Proof. If $a^{-1}|x| \le \|x\| \le b|x|$, then $b^{-1}|x| \le \|x\|_* \le a|x|$, thus by Theorem 6.9,

$$k(X)k(X^*) \ge cn^2\left(\frac{M_X}{b}\right)^2\left(\frac{M_{X^*}}{a}\right)^2 = \frac{cn^2}{(ab)^2}(M_X M_{X^*})^2.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} M_X M_{X^*} = \int_{S^{n-1}} \|x\|\mathrm{d}\sigma(x) \int_{S^{n-1}} \|x\|_*\mathrm{d}\sigma(x) &\ge \left(\int_{S^{n-1}} \sqrt{\|x\| \cdot \|x\|_*}\mathrm{d}\sigma(x)\right)^2 \\ &\ge \left(\int_{S^{n-1}} \sqrt{\langle x, x\rangle}\, \mathrm{d}\sigma(x)\right)^2 = 1, \end{aligned}$$

so

$$k(X)k(X^*) \geq \frac{cn^2}{(ab)^2}.$$

In John's position $ab \leq \sqrt{n}$ (see Theorem 5.8), which finishes the proof. $\square$

6.12 Lemma. *If P is a k-dimensional polytope with f facets such that $B_2^k \subset P \subset aB_2^k$, then $f \geq e^{\frac{k}{2a^2}}$.*

Proof. Let us write P as the intersection of half-spaces $S_j = \{x \in \mathbb{R}^k,\ \langle x, v_j \rangle \leq 1\}$ for some nonzero vectors v_j, $j \leq f$. Since $P \subset aB_2^k$, the union of the caps $S_j^c \cap aS^{k-1} = \{x \in S^{k-1},\ \langle x, v_j \rangle \geq 1\}$ covers the sphere aS^{k-1}, thus rescalling, $S^{k-1} \subset \bigcup_{j \leq f} \frac{1}{a} S_j^c$ and we get from the union bound,

$$1 \leq \sum_{j \leq f} \sigma \left\{ x \in \mathbb{R}^k,\ \langle x, v_j \rangle \geq \frac{1}{a} \right\}$$

Since $B_2^k \subset P$, $|v_j| = \left\langle \frac{v_j}{|v_j|}, v_j \right\rangle \leq 1$, so

$$\left\{ x \in \mathbb{R}^k,\ \langle x, v_j \rangle \geq \frac{1}{a} \right\} \subset \left\{ x \in \mathbb{R}^k,\ \left\langle x, \frac{v_j}{|v_j|} \right\rangle \geq \frac{1}{a} \right\}$$

and by Lemma B.1,

$$1 \leq \sum_{j \leq f} \sigma \left\{ x \in \mathbb{R}^k,\ \left\langle x, \frac{v_j}{|v_j|} \right\rangle \geq \frac{1}{a} \right\} \leq \sum_{j \leq f} e^{-\frac{k}{2a^2}} = fe^{-\frac{k}{2a^2}}.$$

$\square$

Proof of Theorem 6.10 from Theorem 6.11. Let P be an n-dimensional symmetric polytope in $\mathbb{R}^n$ with $f(P)$ facets and $v(P)$ vertices. Consider its norm $\|\cdot\| = \|\cdot\|_P$ and $X = (\mathbb{R}^n, \|\cdot\|)$. Let $k = k(X)$. Then there is a k-dimensional subspace F such that $\frac{1}{2}M_P(B_2^n \cap F) \subset P \cap F \subset 2M_P(B_2^n \cap F)$. Applying Lemma 6.12 to the k-dimensional polytope $P \cap F$ which has at most $f(P)$ facets (every facet of $P \cap F$ comes from a unique facet of P), we get

$$\log f(P) \geq \frac{k}{2 \cdot 4^2} = \frac{1}{32} k(P).$$

Thus by Theorem 6.11,

$$cn \leq k(P)k(P^\circ) \leq 32^2 \log f(P) \log f(P^\circ).$$

To finish, note that $f(P^\circ) = v(P)$. $\square$

6.3 Example: ℓ_p^n

Combining Remark 6.8 and Theorem 6.9, we get that for n-dimensional spaces X whose (possibly dilated) unit balls are in John's position, the largest dimension k for which most k-dimensional sections are 2-Euclidean satisfies

$$k(X) \simeq n \left(\frac{M}{b} \right)^2.$$

Here $a \simeq b$, if there are universal constants c and C such that $ca \le b \le Ca$. Particularly, this gives (up to universal constants) the value of the critical dimension for ℓ_p^n spaces, which we shall now evaluate.

Recall that for a standard Gaussian random vector G in $\mathbb{R}^n$ and a random vector θ uniformly distributed on S^{n-1} we have $\mathbb{E}\|G\| \simeq \sqrt{n}\mathbb{E}\|\theta\|$ (see (A.3)). Therefore,

$$n\left(\frac{M}{b}\right)^2 \simeq \left(\frac{\mathbb{E}\|G\|}{b}\right)^2.$$

For the ℓ_p norms, their Lipschitz constants are easy to find: $\|x\|_p \le |x|$, $p > 2$ and $\|x\|_p \le n^{\frac{1}{p}-\frac{1}{2}}|x|$, $1 < p < 2$ and these are sharp, so

$$b(\ell_p^n) = \begin{cases} n^{\frac{1}{p}-\frac{1}{2}}, & 1 \le p < 2, \\ 1, & p \ge 2. \end{cases}$$

It remains to find ℓ_p norms of a standard Guassian vector.

6.13 Lemma. *Let G be a standard Gaussian random vector in $\mathbb{R}^n$. Then,*

$$\mathbb{E}\|G\|_p \simeq \begin{cases} n^{1/p}\sqrt{p}, & 1 \le p < \log n, \\ \sqrt{\log n}, & p \ge \log n. \end{cases}$$

Proof. We shall write $G = (G_1, \dots, G_n)$. Recall that $(\mathbb{E}|G_1|^p)^{1/p} \simeq \sqrt{p}$.

When $p \ge \log n$, we have the equivalence of the ℓ_p norm and ℓ_∞ norm, $\|x\|_p \simeq \|x\|_\infty$, $x \in \mathbb{R}^n$. As essentially established in the proof of Lemma 6.6, $\mathbb{E}\max_{i\le n}|G_i| \ge c\sqrt{\log n}$. There is a matching upper-bound $\mathbb{E}\max_{i\le n}|G_i| \le C\sqrt{\log n}$, which follows from the union bound (so it does not even use the independence of the G_i). Therefore,

$$\mathbb{E}\|G\|_p \simeq \mathbb{E}\|G\|_\infty \simeq \sqrt{\log n}.$$

Let $p < \log n$. There is an easy upper bound, based on convexity

$$\mathbb{E}\|G\|_p = \mathbb{E}\left(\sum_{i=1}^n |G_i|^p\right)^{1/p} \le \left(\mathbb{E}\sum_{i=1}^n |G_i|^p\right)^{1/p} = n^{1/p}(\mathbb{E}|G_1|^p)^{1/p} \le Cn^{1/p}\sqrt{p}$$

(we have not used that $p < \log n$ here). To reverse this bound, partition $\{1, \dots, n\}$ into roughly $\frac{n}{ce^p}$ subsets I_j of roughly equal size exceeding ce^p. Then,

$$\begin{aligned}\mathbb{E}\|G\|_p = \mathbb{E}\left(\sum_j \sum_{i\in I_j} |G_i|^p\right)^{1/p} &\ge \mathbb{E}\left(\sum_j \left(\max_{i\in I_j}|G_i|\right)^p\right)^{1/p} \\ &\ge \left(\sum_j \left(\mathbb{E}\max_{i\in I_j}|G_i|\right)^p\right)^{1/p} \\ &\ge c\left(\frac{n}{ce^p}\left(c\sqrt{\log|I_j|}\right)^p\right)^{1/p} \ge cn^{1/p}\sqrt{p}.\end{aligned}$$

□

6.14 Corollary. *The critical dimension of n-dimensional ℓ_p spaces up to universal constants equals*

$$k(\ell_p^n) \simeq \begin{cases} n, & 1 \le p < 2, \\ pn^{2/p}, & 2 \le p < \log n, \\ \log n, & p \ge \log n. \end{cases}$$

6.4 Proportional dimension

We remark that when $1 \le p < 2$, Corollary 6.14 says that B_p^n has critical dimension $\simeq n$, that is it has Euclidean sections of proportional dimension (this is not so surprising given that it has 2^n facets). The maximal volume ellipsoid of B_p^n is $n^{\frac{1}{2}-\frac{1}{p}} B_2^n$ (reason: $n^{-1/p}(\pm 1, \ldots, \pm 1)$ are contact points which clearly give the decomposition of the identity). Thus the volume ratio equals

$$\mathrm{vr}(B_p^n) = \left(\frac{|B_p^n|}{|n^{1/2-1/p} B_2^n|} \right)^{1/n} \simeq \text{const}, \qquad 1 \le p < 2.$$

This is not accidental as explained in the next theorem.

6.15 Theorem. *Let K be a symmetric convex body in $\mathbb{R}^n$ such that $B_2^n \subset K$ and $|K| = \alpha^n |B_2^n|$, $\alpha > 1$. Then for every $1 \le k \le n$, we have that there is a subset Γ of k-dimensional subspaces of Haar measure $\nu(\Gamma) \ge 1 - e^{-n}$ such that*

$$\forall F \in \Gamma \qquad B_2^n \cap F \ \subset \ K \cap F \ \subset \ (8e\alpha)^{\frac{n}{n-k}} (B_2^n \cap F).$$

In particular, if α is a constant, k is roughly cn, we get that K has k-dimensional C-Euclidean sections.

Proof. Let $\| \cdot \| = \| \cdot \|_K$ be the norm given by K. We want to find subspaces F such that $(C\alpha)^{-\frac{n}{n-k}} \le \|x\| \le 1$ for $x \in S^{n-1} \cap F = S_F$. Since $B_2^n \subset K$, $\|x\| \le |x|$, so the upper bound is clear. To go about the lower bound, note that by the factorisation of Haar measures (A.6) and the volume formula in polar coordinates,

$$\int_{G_{n,k}} \int_{S_F} \|x\|^{-n} \mathrm{d}\sigma_F(x) \mathrm{d}\nu_{n,k}(F) = \int_{S^{n-1}} \|x\|^{-n} \mathrm{d}\sigma(x) = \frac{|K|}{|B_2^n|} = \alpha^n.$$

This gives that for

$$\Gamma = \{F, \int_{S_F} \|x\|^{-n} \mathrm{d}\sigma_F(x) \le (e\alpha)^n\},$$

by Chebyshev's inequality we have

$$\nu_{n,k}(\Gamma^c) \le e^{-n}.$$

Fix $F \in \Gamma$. Our goal is to show $\|x\| \ge (C\alpha)^{-\frac{n}{n-k}}$, $x \in S_F$. By Chebyshev's inequality and the definition of Γ,

$$(e\alpha)^n \ge \int_{S_F} \|x\|^{-n} \mathbf{1}_{\{\|x\| < r\}} \mathrm{d}\sigma_F(x) \ge r^{-n} \sigma_F \left\{ x \in S_F, \ \|x\| < r \right\},$$

thus for $A = \{x \in S_F,\ \|x\| \geq r\}$, $\sigma_F(A) \geq 1 - (re\alpha)^n$. Fix $x \in S_F$ and consider a spherical cap $C(x)$ around x or radius $r/2$. By our lower bound for the measure of spherical caps, $\sigma(C(x)) \geq \left(\frac{r}{8}\right)^k$ (Theorem B.2). Consequently,

$$\begin{aligned}\sigma_F(A \cap C(x)) = \sigma_F(A) + \sigma_F(C(x)) - \sigma_F(A \cup C(x)) &\geq 1 - (re\alpha)^n + \left(\frac{r}{8}\right)^k - 1 \\ &= \left(\frac{r}{8}\right)^k - (re\alpha)^n.\end{aligned}$$

For any r such that this measure is positive, that is $r < r_0$ with $r_0 = 8^{-\frac{k}{n-k}}(e\alpha)^{-\frac{n}{n-k}}$, for $y \in A \cap C(x)$ we get

$$\|x\| \geq \|y\| - \|x - y\| \geq r - |x - y| \geq \frac{r}{2}.$$

Since $\frac{r_0}{2} = 2^{-1}8^{-\frac{k}{n-k}}(e\alpha)^{-\frac{n}{n-k}} > 8^{-\frac{n}{n-k}}(e\alpha)^{-\frac{n}{n-k}}$, $k < n$, by taking r appropriately close to r_0, we get

$$\|x\| \geq (8e\alpha)^{-\frac{n}{n-k}}.$$

□

Now we prove a *global* version of the previous theorem.

6.16 Theorem. *Let K be a symmetric convex body in $\mathbb{R}^n$ such that $B_2^n \subset K$ and $|K| = \alpha^n |B_2^n|$, $\alpha > 1$. Then there exist an orthogonal map $U \in O(n)$ such that*

$$B_2^n \ \subset\ K \cap UK \ \subset\ 16\alpha^2 B_2^n.$$

Proof. Let $\|\cdot\| = \|\cdot\|_K$ be the norm given by K. Since $B_2^n \subset K$, for any $U \in O(n)$, $B_2^n \subset UK$ and consequently $B_2^n \subset K \cap UK$. It suffices to find U such that $K \cap UK \subset C\alpha^2 B_2^n$, or in other words $\|x\|_{K\cap UK} \geq \frac{1}{C\alpha^2}$ for all $x \in S^{n-1}$. We have

$$\|x\|_{K\cap UK} = \max\{\|x\|, \|x\|_{UK}\} = \max\{\|x\|, \|U^T x\|\} \geq \frac{\|x\| + \|U^T x\|}{2}.$$

Let

$$N_U(x) = \frac{\|x\| + \|U^T x\|}{2}.$$

Computing its appropriate average yields

$$\begin{aligned}\int_{O(n)}\int_{S^{n-1}} N_U(\theta)^{-2n}\mathrm{d}\sigma(\theta)\mathrm{d}\nu(U) &\leq \int_{O(n)}\int_{S^{n-1}} \frac{1}{\|\theta\|^n\|U^T\theta\|^n}\mathrm{d}\sigma(\theta)\mathrm{d}\nu(U) \\ &= \int_{S^{n-1}} \frac{1}{\|\theta\|^n}\left(\int_{O(n)} \frac{1}{\|U^T\theta\|^n}\mathrm{d}\nu(U)\right)\mathrm{d}\sigma(\theta) \\ &= \int_{S^{n-1}}\int_{S^{n-1}} \frac{1}{\|\theta\|^n\|\theta'\|^n}\mathrm{d}\sigma(\theta)\mathrm{d}\sigma(\theta') \\ &= \left(\int_{S^{n-1}} \frac{1}{\|\theta\|^n}\mathrm{d}\sigma(\theta)\right)^2 = \left(\frac{|K|}{|B_2^n|}\right)^2 = \alpha^{2n},\end{aligned}$$

where we have used (A.4). Therefore, there exist $U \in O(n)$ such that

$$\int_{S^{n-1}} N_U(\theta)^{-2n} \mathrm{d}\sigma(\theta) \le \alpha^{2n}.$$

Let ε be the minimum $\min_{\theta \in S^{n-1}} N_U(\theta)$ attained at, say $\theta_0 \in S^{n-1}$. Since $\|\cdot\|$ is 1-Lipschitz (because $B_2^n \subset K$), for any $\theta \in S^{n-1}$ such that $|\theta - \theta_0| < \varepsilon$, we get $N_U(\theta) \le N_U(\theta - \theta_0) + N_U(\theta_0) \le 2\varepsilon$. Thus,

$$\begin{aligned}\alpha^{2n} &\ge \int_{S^{n-1}} N_U(\theta)^{-2n} \mathrm{d}\sigma(\theta) \ge \int_{S^{n-1}} N_U(\theta)^{-2n} \mathbf{1}_{\{|\theta-\theta_0|<\varepsilon\}} \mathrm{d}\sigma(\theta) \\ &\ge (2\varepsilon)^{-2n} \sigma\{\theta,\ |\theta - \theta_0| < \varepsilon\} \ge (2\varepsilon)^{-2n} \left(\frac{\varepsilon}{4}\right)^n = \frac{1}{(16\varepsilon)^n},\end{aligned}$$

where we have used again our lower bound for the measure of spherical caps – Theorem B.2. This finishes the proof because

$$\min_\theta N_U(\theta) = \varepsilon \ge \frac{1}{16\alpha^2}.$$

□

When applied to B_1^n, the above proof gives Kashin's theorem about n-dimensional Euclidean sections of B_1^{2n}.

6.17 Theorem (Kashin)**.** *There are two orthonormal bases $(y_i)_{i=1}^n$ and $(y_i)_{i=n+1}^{2n}$ of $\mathbb{R}^n$ such that*

$$\frac{1}{18}\sqrt{n}|x| \le \sum_{i=1}^{2n} |\langle x, y_i \rangle| \le 2\sqrt{n}|x|, \qquad x \in \mathbb{R}^n.$$

Proof. We have $B_2^n \subset \sqrt{n} B_1^n$ and it can be checked that for $n \ge 1$,

$$\frac{|\sqrt{n}B_1^n|}{|B_2^n|} = \frac{n^{n/2}2^n/n!}{\sqrt{\pi}^n/\Gamma(1+n/2)} \le \left(\frac{3}{2}\right)^n.$$

Taking $\alpha = \frac{3}{2}$, the proof of Theorem 6.16 gives an orthogonal map $U \in O(n)$ such that

$$\frac{1}{16\alpha^2}|x| \le \frac{1}{\sqrt{n}} \frac{\|x\|_1 + \|Ux\|_1}{2} \le |x|,$$

so setting $y_i = e_i$, $y_{n+i} = U^T e_i$, $i \le n$ gives the result. □

7 Distribution of mass

In this chapter we show two general results concerning distribution of mass in convex bodies: 1) large deviations inequality for the Euclidean norm due to Paouris, and 2) small ball estimates for arbitrary norms due to Latała. Instead of just uniform distributions on convex bodies, we shall consider log-concave distributions.

7.1 Large deviations bound

Paouris established the following general large deviations inequality for the Euclidean norm.

7.1 Theorem. *There is a universal constant $C > 0$ such that for every isotropic log-concave random vector X in $\mathbb{R}^n$, we have*

$$\mathbb{P}\left(|X| \geq Ct\sqrt{n}\right) \leq e^{-t\sqrt{n}}, \qquad t \geq 1. \tag{7.1}$$

Instead of Paouris' original approach, we shall follow a later argument which goes through a moment comparison inequality.

7.2 Theorem. *There is a universal constant $C > 0$ such that for a log-concave random vector X in $\mathbb{R}^n$ and $q \geq 1$, we have*

$$(\mathbb{E}|X|^q)^{1/q} \leq C\Big(\mathbb{E}|X| + \sup_{\theta \in S^{n-1}} (\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q}\Big). \tag{7.2}$$

Proof that Theorem 7.2 implies Theorem 7.1. Let X be an isotropic log-concave random vector in $\mathbb{R}^n$. By isotropicity, $\mathbb{E}|\langle X, \theta\rangle|^2 = 1$ for every $\theta \in S^{n-1}$ and $\mathbb{E}|X|^2 = n$. Thus, $\mathbb{E}|X| \leq (\mathbb{E}|X|^2)^{1/2} = \sqrt{n}$. Moreover, by the moment comparison inequality for semi-norms from Theorem 3.15, $(\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q} \leq Cq(\mathbb{E}|\langle X, \theta\rangle|^2)^{1/2} = Cq$, for $q \geq 2$. Note this bound remains true for $q \geq 1$ because for $q \leq 2$, $(\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q} \leq (\mathbb{E}|\langle X, \theta\rangle|^2)^{1/2} = 1$. Therefore, (7.2) yields for every $q \geq 1$,

$$(\mathbb{E}|X|^q)^{1/q} \leq C(\sqrt{n} + q).$$

By Chebyshev's inequality,

$$\mathbb{P}\left(|X| \geq C't\sqrt{n}\right) \leq (C't\sqrt{n})^{-q}\mathbb{E}|X|^q \leq \left(\frac{C}{C'}\right)^q (t\sqrt{n})^{-q}(\sqrt{n} + q)^q.$$

Given $t \geq 1$, let $q = t\sqrt{n}$. Let $C' = 2Ce$. Then the above becomes

$$\mathbb{P}\left(|X| \geq C't\sqrt{n}\right) \leq \left(\frac{1}{2e}\right)^q \left(\frac{1+t}{t}\right)^q \leq \left(\frac{1}{e}\right)^q = e^{-t\sqrt{n}}.$$

□

To prove Theorem 7.2, we need two technical lemmas.

7.3 Lemma. *For a random vector X in $\mathbb{R}^n$, a norm $\|\cdot\|$ on $\mathbb{R}^n$ and $q \geq 1$, we have*

$$\inf_{\theta \in S^{n-1}} (\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q} \leq \frac{(\mathbb{E}\|X\|^q)^{1/q}}{\mathbb{E}\|X\|}\mathbb{E}|X|.$$

Proof. Let K be the unit ball with respect to $\|\cdot\|$ and let b be the smallest number such that $B_2^n \subset bK$ (the Lipschitz constant of $\|\cdot\|$). Let θ_0 be a contact point of B_2^n and bK. Then $\|\theta_0\|_{bK} = 1 = \|\theta_0\|_{(bK)^\circ} = b\|\theta_0\|_{K^\circ}$ (recall Remark 5.4). Thus,

$$|\langle \theta_0, x\rangle| \leq \|\theta_0\|_{K^\circ}\|x\| = b^{-1}\|x\|, \qquad x \in \mathbb{R}^n,$$

which gives

$$(\mathbb{E}|\langle X, \theta_0\rangle|^q)^{1/q} \leq b^{-1}(\mathbb{E}\|X\|^q)^{1/q}.$$

By the definition of b, $\|x\| \leq b|x|$, so $\mathbb{E}\|X\| \leq b\mathbb{E}|X|$, which combined with the above finishes the proof. □

7.4 Lemma. *For a symmetric log-concave random vector X in $\mathbb{R}^n$, there is a norm $\|\cdot\|$ on $\mathbb{R}^n$ such that*

$$(\mathbb{E}\|X\|^n)^{1/n} \leq 500\mathbb{E}\|X\|.$$

Proof. Without loss of generality assume that the support of X is n-dimensional and consider its log-concave density $f : \mathbb{R}^n \to [0, \infty)$ which is even. Since f is even and log-concave, $f(0) = \|f\|_\infty$. Let

$$K = \{x \in \mathbb{R}^n, \ f(x) \geq 25^{-n} f(0)\}$$

and let $\|\cdot\|$ be the norm whose unit ball is K.

First, we bound $\mathbb{E}\|X\|$ below. Note that

$$1 \geq \int_K f \geq 25^{-n} f(0)|K|$$

and

$$\mathbb{P}\left(\|X\| \leq \frac{1}{50}\right) = \int_{\frac{1}{50}K} f \leq |\frac{1}{50}K| \cdot f(0) = \frac{1}{50^n}|K|f(0).$$

Using the previous estimate yields

$$\mathbb{P}\left(\|X\| \leq \frac{1}{50}\right) \leq \frac{1}{50^n}25^n = \frac{1}{2^n} \leq \frac{1}{2}.$$

By Chebyshev's inequality we can conclude that

$$\mathbb{E}\|X\| \geq \frac{1}{50}\mathbb{P}\left(\|X\| > \frac{1}{50}\right) \geq \frac{1}{100}.$$

Second, we bound $\mathbb{E}\|X\|^n$ above. Note that

$$\mathbb{E}\|X\|^n = \mathbb{E}\|X\|^n \mathbf{1}_{\{\|X\| \leq 1\}} + \mathbb{E}\|X\|^n \mathbf{1}_{\{\|X\| > 1\}} \leq 1 + \mathbb{E}\|X\|^n \mathbf{1}_{K^c}.$$

On K^c we have $f(0) > 25^n f(x)$ and thus

$$f_{2X}(x) = 2^{-n} f\left(\frac{x}{2}\right) \geq 2^{-n}\sqrt{f(x)f(0)} \geq \left(\frac{5}{2}\right)^n f(x),$$

so

$$\begin{aligned}\mathbb{E}\|X\|^n \mathbf{1}_{K^c} = \int_{K^c} \|x\|^n f(x) \leq \left(\frac{2}{5}\right)^n \int_{K^c} \|x\|^n f_{2X}(x) &\leq \left(\frac{2}{5}\right)^n \mathbb{E}\|2X\|^n \\ &= \left(\frac{4}{5}\right)^n \mathbb{E}\|X\|^n.\end{aligned}$$

Using this in the second last inequality, we obtain

$$\mathbb{E}\|X\|^n \leq \frac{1}{1-\left(\frac{4}{5}\right)^n} \leq 5.$$

□

Proof of Theorem 7.2. Without loss of generality we can assume that X is symmetric. Otherwise, consider X and take its independent copy X'. Then $X - X'$ is a symmetric log-concave random vector, so knowing the theorem for such vectors gives an upper bound for $\mathbb{E}|X - X'|^q$,

$$\begin{aligned}(\mathbb{E}|X-X'|^q)^{1/q} &\leq C\Big(\mathbb{E}|X-X'| + \sup_{\theta\in S^{n-1}} (\mathbb{E}|\langle X-X',\theta\rangle|^q)^{1/q}\Big) \\ &\leq 2C\Big(\mathbb{E}|X| + \sup_{\theta\in S^{n-1}} (\mathbb{E}|\langle X,\theta\rangle|^q)^{1/q}\Big).\end{aligned}$$

By the triangle inequality $|X| \leq |X - \mathbb{E}X| + |\mathbb{E}X|$, so by the triangle inequality in L_q,

$$(\mathbb{E}|X|^q)^{1/q} \leq (\mathbb{E}|X-\mathbb{E}X|^q)^{1/q} + |\mathbb{E}X|.$$

By Jensen's inequality, $\mathbb{E}|X - \mathbb{E}X|^q = \mathbb{E}|X - \mathbb{E}X'|^q \leq \mathbb{E}|X - X'|^q$ and $|\mathbb{E}X| \leq \mathbb{E}|X|$. Combined with the previous estimates, this gives

$$(\mathbb{E}|X|^q)^{1/q} \leq C'\Big(\mathbb{E}|X| + \sup_{\theta\in S^{n-1}} (\mathbb{E}|\langle X,\theta\rangle|^q)^{1/q}\Big).$$

Assume from now on that X is symmetric. Define a function $h : \mathbb{R}^n \to [0,\infty)$,

$$h(u) = (\mathbb{E}|\langle X,u\rangle|^q)^{1/q}, \qquad u \in \mathbb{R}^n,$$

which is a semi-norm. Let G be a standard Gaussian random vector in $\mathbb{R}^n$. Conditioned on the value of X, $\langle X, G\rangle$ has the same distribution as $|X|G_1$, hence we have

$$\mathbb{E}h(G)^q = \mathbb{E}|\langle X,G\rangle|^q = \mathbb{E}|X|^q|G_1|^q = \mathbb{E}|G_1|^q\mathbb{E}|X|^q,$$

that is, introducing

$$c_q = (\mathbb{E}|G_1|^q)^{1/q} = \Theta(\sqrt{q}),$$

we have

$$(\mathbb{E}|X|^q)^{1/q} = \frac{1}{c_q}(\mathbb{E}h(G)^q)^{1/q}.$$

Let b be the Lipschitz constant of h, that is

$$b = \sup_{\theta \in S^{n-1}} h(\theta) = \sup_{\theta \in S^{n-1}} (\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q}.$$

By the Gaussian concentration inequality for Lipschitz functions (see Corollary 3.5 and Remark 3.6),

$$\mathbb{P}\left(|h(G) - \mathbb{E}h(G)| \geq s\right) \leq Ce^{-cs^2/b^2}.$$

This and a standard computation of moments using tails yields

$$\mathbb{E}|h(G) - \mathbb{E}h(G)|^q = \int_0^\infty qs^{q-1}\mathbb{P}\left(|h(G) - \mathbb{E}h(G)| \geq s\right) \mathrm{d}s \leq Cc_q^q b^q.$$

By the triangle inequality in L_q, we can write

$$(\mathbb{E}h(G)^q)^{1/q} \leq (\mathbb{E}|h(G) - \mathbb{E}h(G)|^q)^{1/q} + \mathbb{E}h(G).$$

Putting everything together,

$$\begin{aligned}(\mathbb{E}|X|^q)^{1/q} = \frac{1}{c_q}(\mathbb{E}h(G)^q)^{1/q} &\leq \frac{1}{c_q}\Big(\mathbb{E}h(G) + Cc_q b\Big) \\ &\leq C\Big(\frac{1}{\sqrt{q}}\mathbb{E}h(G) + \sup_{\theta \in S^{n-1}} (\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q}\Big).\end{aligned}$$

To show (7.2), it thus remains to show that $\frac{1}{\sqrt{q}}\mathbb{E}h(G)$ is upper bounded (up to a constant) by either $\mathbb{E}|X|$ or b or their sum.

Case 1. If $q \geq c\left(\frac{\mathbb{E}h(G)}{b}\right)^2$, then $\frac{1}{\sqrt{q}}\mathbb{E}h(G) \leq \frac{1}{\sqrt{c}}b$, so there is nothing to do in this case.

Case 2. If $q \leq c\left(\frac{\mathbb{E}h(G)}{b}\right)^2$, then by Dvoretzky's theorem 6.4 with $\varepsilon = \frac{1}{2}$ applied to h (note that $M_h = \int_{S^{n-1}} h = (1 + o(1))\frac{\mathbb{E}h(G)}{\sqrt{n}}$), there is subset Γ of subspaces F of $\mathbb{R}^n$ of dimension k, say $q \leq k < 2q$ such that

$$\frac{2}{3}\frac{\mathbb{E}h(G)}{\sqrt{n}}|x| \leq h(x) \leq \frac{3}{2}\frac{\mathbb{E}h(G)}{\sqrt{n}}|x|, \qquad x \in F$$

and

$$\nu_{n,k}(\Gamma) \geq 1 - e^{-c'\left(\frac{\mathbb{E}h(G)}{b}\right)^2} \geq 1 - e^{-c'q/c} \geq 1 - e^{-c'/c} > \frac{2}{3}.$$

Let P_F be the projection onto F and $S_F = S^{n-1} \cap F$. By Lemmas 7.3 and 7.4 applied to $Y = P_F(X)$,

$$\begin{aligned}\inf_{\theta \in S^{n-1}} (\mathbb{E}|\langle X, \theta\rangle|^q)^{1/q} \leq \inf_{\theta \in S_F} (\mathbb{E}|\langle Y, \theta\rangle|^q)^{1/q} &\leq \inf_{\theta \in S_F} (\mathbb{E}|\langle Y, \theta\rangle|^k)^{1/k} \\ &\leq \frac{(\mathbb{E}\|Y\|^k)^{1/k}}{\mathbb{E}\|Y\|}\mathbb{E}|Y| \\ &\leq 500\mathbb{E}|Y| = 500\mathbb{E}|P_F(X)|.\end{aligned}$$

Since for every $F \in \Gamma$ and $\theta \in S_F$,

$$\mathbb{E}h(G) \leq \frac{3}{2}\sqrt{n}h(\theta),$$

by taking infimum over θ and using the above,

$$\mathbb{E}h(G) \le \frac{3}{2}\sqrt{n}500\mathbb{E}|P_F(X)| = 750\sqrt{n}\mathbb{E}|P_F(X)|.$$

To finish, note that for every $x \in \mathbb{R}^n$, we have

$$\int_{G_{n,k}} |P_E(x)|^2 \mathrm{d}\nu_{n,k}(E) = \frac{k}{n}|x|^2.$$

Explanation: we can treat E as the image of $E_0 = \mathrm{span}\{e_1, \ldots, e_k\}$ under a uniform random orthogonal matrix U; then $|P_E(x)|^2 = \sum_{i=1}^k |\langle Ue_i, x\rangle|^2$, $\langle Ue_i, x\rangle$ has the same distribution as $\eta_i|x|$, where η is uniformly distributed on S^{n-1} and $\mathbb{E}|\eta_i|^2 = \frac{1}{n}$. Thus

$$\int_{G_{n,k}} |P_E(x)| \mathrm{d}\nu_{n,k}(E) \le \left(\int_{G_{n,k}} |P_E(x)|^2 \mathrm{d}\nu_{n,k}(E)\right)^{1/2} = \sqrt{\frac{k}{n}}|x|.$$

By Chebyshev's inequality,

$$\nu_{n,k}\left\{F \in G_{n,k},\ \mathbb{E}|P_F(X)| \le C\sqrt{\frac{k}{n}}\mathbb{E}|X|\right\} \ge 1 - \frac{1}{C}.$$

Choosing, say $C = 3$, there is nontrivial intersection of this even with Γ and picking a subspace F belonging to both we finally get

$$\mathbb{E}h(G) \le 750\sqrt{n}\mathbb{E}|P_F(X)| \le 750\sqrt{n} \cdot 3\sqrt{\frac{k}{n}}\mathbb{E}|X| \le C\sqrt{q}\mathbb{E}|X|.$$

□

7.2 Small ball estimates

The goal of this section is to show Latała's inequality.

7.5 Theorem. *For every log-concave random vector X in $\mathbb{R}^n$ and every norm $\|\cdot\|$ on $\mathbb{R}^n$, we have*

$$\mathbb{P}\left(\|X\| < t\mathbb{E}\|X\|\right) \le 384t, \qquad t \in [0,1]. \tag{7.3}$$

Proof. Let K be the unit ball with respect to $\|\cdot\|$. Without loss of generality we can assume that X has a density on $\mathbb{R}^n$ (otherwise, consider $X + \varepsilon Y$ for an independent Y being uniform on the unit ball K of $\|\cdot\|$ and then use

$$\mathbb{P}\left(\|X\| < t\mathbb{E}\|X\|\right) = \lim_{\varepsilon\to 0}\mathbb{P}\left(\|X\| + \varepsilon < t\mathbb{E}\|X\|\right)$$

as well as

$$\mathbb{P}\left(\|X\| + \varepsilon < t\mathbb{E}\|X\|\right) \le \mathbb{P}\left(\|X\| + \varepsilon\|Y\| < t\mathbb{E}\|X\|\right) \le \mathbb{P}\left(\|X + \varepsilon Y\| < t\mathbb{E}\|X\|\right)$$

and $\mathbb{E}\|X\| \le \mathbb{E}\|X + \varepsilon Y\|$). This guarantees that $\|X\|$ has no atoms and its distribution function is thus continuous. Choose then $\alpha > 0$ to be the smallest number such that

$$\mathbb{P}\left(\|X\| \le \alpha\right) = \frac{2}{3}.$$

By Borell's lemma (Theorem 3.12 and Remark 3.14), we have

$$\mathbb{P}\left(\|X\| > t\alpha\right) \le \frac{2}{3}\left(\frac{1-\frac{2}{3}}{\frac{2}{3}}\right)^{\frac{t+1}{2}} = \frac{2}{3}\left(\frac{1}{2}\right)^{\frac{t+1}{2}}, \qquad t \ge 1.$$

In particular,

$$\mathbb{P}\left(\|X\| > 3\alpha\right) \le \frac{1}{6},$$

thus

$$\mathbb{P}\left(\alpha < \|X\| \le 3\alpha\right) = \mathbb{P}\left(\|X\| \le 3\alpha\right) - \mathbb{P}\left(\|X\| \le \alpha\right) \ge \frac{5}{6} - \frac{2}{3} = \frac{1}{6}.$$

Fix $k \ge 1$. Define the rings

$$R(u) = \left\{x \in \mathbb{R}^n,\ u - \frac{\alpha}{2k} < \|x\| \le u + \frac{\alpha}{2k}\right\}, \qquad u \ge \frac{\alpha}{2k}.$$

Since

$$\{x \in \mathbb{R}^n,\ \alpha \le \|x\| < 3\alpha\} = \bigcup_{j=1}^{2k} R\left(\alpha + \frac{2j-1}{2k}\alpha\right),$$

for $u_0 = \alpha + \frac{2j_0 - 1}{2k}\alpha$ for some $1 \le j_0 \le 2k$, we have

$$\mathbb{P}\left(X \in R(u_0)\right) \ge \frac{1}{12k}.$$

Note that for every $0 \le \lambda \le 1$ and $u \ge \frac{\alpha}{2k}$, we have

$$\lambda R(u) + (1-\lambda)\frac{\alpha}{2k}K \subset R(\lambda u).$$

Indeed, if $x \in R(u)$ and $y \in K$, then

$$\left\|\lambda x + (1-\lambda)\frac{\alpha}{2k}y\right\| \le \lambda\|x\| + (1-\lambda)\frac{\alpha}{2k}\|y\| \le \lambda\left(u + \frac{\alpha}{2k}\right) + (1-\lambda)\frac{\alpha}{2k} = \lambda u + \frac{\alpha}{2k}$$

and

$$\left\|\lambda x + (1-\lambda)\frac{\alpha}{2k}y\right\| \ge \lambda\|x\| - (1-\lambda)\frac{\alpha}{2k}\|y\| \ge \lambda\left(u - \frac{\alpha}{2k}\right) - (1-\lambda)\frac{\alpha}{2k} = \lambda u - \frac{\alpha}{2k}.$$

Claim. $\mathbb{P}\left(\|X\| \le \frac{\alpha}{2k}\right) \le \frac{48}{k}$, $k = 1, 2, \ldots$.

Proof of the claim. Suppose it does not hold, so there is $k_0 \ge 1$ such that

$$\mathbb{P}\left(\|X\| \le \frac{\alpha}{2k_0}\right) > \frac{48}{k_0}.$$

As explained earlier, for this k_0, there is $u_0 > \alpha$ of the form $\alpha + \frac{2j_0-1}{2k_0}\alpha$ such that

$$\mathbb{P}\left(X \in R(u_0)\right) \ge \frac{1}{12k_0}.$$

By log-concavity,

$$\begin{aligned}\mathbb{P}\left(X \in R(\lambda u_0)\right) &\ge \mathbb{P}\left(X \in \lambda R(u_0) + (1-\lambda)\frac{\alpha}{2k_0}K\right)\\ &\ge \mathbb{P}\left(X \in R(u_0)\right)^{\lambda}\,\mathbb{P}\left(\|X\| \le \frac{\alpha}{2k_0}\right)^{1-\lambda}\\ &\ge \left(\frac{1}{12k_0}\right)^{\lambda}\left(\frac{48}{k_0}\right)^{1-\lambda}.\end{aligned}$$

Note that for $\lambda \le \frac{1}{2}$, $\frac{48^{1-\lambda}}{12^\lambda} = \frac{48}{(48\cdot 12)^\lambda} \ge \frac{48}{\sqrt{48\cdot 12}} = 2$, so for every $u \le \frac{u_0}{2}$,

$$\mathbb{P}\left(X \in R(u)\right) = \mathbb{P}\left(X \in R(\lambda u_0)\right) \ge \frac{2}{k_0}.$$

Consider the sets $A_j = R(\frac{j}{k_0}\alpha)$, $1 \le j \le \frac{k_0}{2}$. They are disjoint. Since $\frac{j}{k_0}\alpha \le \frac{\alpha}{2} < \frac{u_0}{2}$, $\mathbb{P}\left(X \in A_j\right) \ge \frac{2}{k_0}$, so $\mathbb{P}\left(X \in \bigcup A_j\right) \ge \lfloor \frac{k_0}{2} \rfloor \cdot \frac{2}{k_0}$. On the other hand, $\bigcup A_j$ is disjoint from $\frac{\alpha}{2k_0}K$ and $\mathbb{P}\left(\|X\| \le \frac{\alpha}{2k_0}\right) > \frac{48}{k_0}$. Thus,

$$\left\lfloor \frac{k_0}{2} \right\rfloor \cdot \frac{2}{k_0} + \frac{48}{k_0} \le 1,$$

which gives a contradiction. □

Let $0 < t \le \frac{1}{2}$. Take an integer $k \ge 1$ such that $\frac{1}{4k} \le t \le \frac{1}{2k}$. Then, by the claim,

$$\mathbb{P}\left(\|X\| \le t\alpha\right) \le \mathbb{P}\left(\|X\| \le \frac{\alpha}{2k}\right) \le \frac{48}{k} \le 4 \cdot 48t.$$

To finish the argument, observe that $\mathbb{E}\|X\|$ is comparable with α. We have,

$$\mathbb{E}\|X\| = \int_0^\infty \mathbb{P}\left(\|X\| > t\right) \mathrm{d}t \le \alpha + \int_\alpha^\infty \mathbb{P}\left(\|X\| > t\right) \mathrm{d}t = \alpha + \alpha \int_1^\infty \mathbb{P}\left(\|X\| > t\alpha\right) \mathrm{d}t.$$

Using Borell's lemma,

$$\int_1^\infty \mathbb{P}\left(\|X\| > t\alpha\right) \mathrm{d}t \le \int_1^\infty \frac{2}{3}\left(\frac{1}{2}\right)^{\frac{t+1}{2}} \mathrm{d}t = \frac{2}{3\log 2} < 1.$$

Hence, $\mathbb{E}\|X\| \le 2\alpha$ and we get

$$\mathbb{P}\left(\|X\| \le t\mathbb{E}\|X\|\right) \le \mathbb{P}\left(\|X\| \le 2t\alpha\right) \cdot 4 \cdot 48 \cdot 2t = 384t,$$

for $t \le \frac{1}{4}$. For $\frac{1}{4} < t \le 1$, trivially,

$$\mathbb{P}\left(\|X\| \le t\mathbb{E}\|X\|\right) \le 1 \le 4t.$$

□

As a corollary, we obtain a moment comparison inequality.

7.6 Theorem. *For every log-concave random vector X in $\mathbb{R}^n$, every norm $\|\cdot\|$ on $\mathbb{R}^n$ and $-1 < q < 0$, we have*

$$\mathbb{E}\|X\| \le \frac{e^{384}}{1+q}\left(\mathbb{E}\|X\|^q\right)^{1/q}. \tag{7.4}$$

Proof. We can assume that $\mathbb{E}\|X\| = 1$. Let $p = -q \in (0,1)$. We have,

$$\mathbb{E}\|X\|^q = \mathbb{E}\left(\frac{1}{\|X\|}\right)^p = \int_0^\infty pt^{p-1}\mathbb{P}\left(\frac{1}{\|X\|} > t\right) \mathrm{d}t \le 1 + \int_1^\infty pt^{p-1}\mathbb{P}\left(\|X\| < \frac{1}{t}\right) \mathrm{d}t.$$

By (7.3),

$$\int_1^\infty pt^{p-1}\mathbb{P}\left(\|X\| < \frac{1}{t}\right) \mathrm{d}t \le 384\int_1^\infty pt^{p-2} \mathrm{d}t = \frac{384p}{1-p}.$$

This gives,

$$\frac{1}{1+q}\left(\mathbb{E}\|X\|^q\right)^{1/q} \geq \frac{1}{1-p}\left(1+\frac{384p}{1-p}\right)^{-1/p} = (1+383p)^{-1/p}(1-p)^{1/p-1}.$$

Clearly, $(1+383p)^{-1/p} \geq e^{383p\frac{-1}{p}} = e^{-383}$. For the second term, we check (by taking the logarithm and differentiating) that $p \to (1-p)^{1/p-1}$ increases on $(0,1)$, so it is lower bounded by its limit at $p \to 0$, which is e^{-1}. □

8 Brascamp-Lieb inequalities

The goal of this section is to present Brascamp-Lieb inequalities and their reverse form, due to Barthe. We shall follow his unified approach to both results. As applications, we give a proof of Ball's inequality (Theorem 5.14) omitted earlier, Young's inequality for convolutions with sharp constants, derive the entropy power inequality and, as a good excuse, discuss the relation between entropy and the slicing problem (Conjecture 4.15).

8.1 Main result

Given $m \geq n$, positive numbers $c_1, \ldots, c_m$ such that $\sum_{i=1}^m c_i = n$ and vectors $v_1, \ldots, v_m$ in $\mathbb{R}^n$ define for integrable functions $f_1, \ldots, f_m : \mathbb{R} \to [0, \infty)$ the following operators

$$J(f_1, \ldots, f_m) = \int_{\mathbb{R}^n} \prod_{i=1}^m f_i(\langle x, v_i \rangle)^{c_i} \mathrm{d}x$$

and

$$I(f_1, \ldots, f_m) = \int_{\mathbb{R}^n}^{\star} \sup \left\{ \prod_{i=1}^m f_i(t_i)^{c_i},\ x = \sum_{i=1}^m c_i t_i v_i \right\} \mathrm{d}x.$$

Here, for a not necessarily measurable function f (as may be the case for the supremum above), we use its outer integral,

$$\int_{\mathbb{R}^n}^{\star} f = \sup \left\{ \int_{\mathbb{R}^n} h,\ h \leq f,\ h \text{ is measurable} \right\}.$$

We are interested in best constants E, F in the following inequalities

$$J(f_1, \ldots, f_m) \leq F \cdot \prod_{i=1}^m \left(\int_{\mathbb{R}} f_i \right)^{c_i}$$

and

$$I(f_1, \ldots, f_m) \geq E \cdot \prod_{i=1}^m \left(\int_{\mathbb{R}} f_i \right)^{c_i}.$$

The main deep result is that these constants come from testing the inequalities with Gaussian functions (note that the inequalities do not change when f_i is replaced with $\lambda_i f_i$ for some $\lambda_i > 0$, thus it suffices to consider centred Gaussian functions).

8.1 Theorem (Brascamp-Lieb inequalities)**.** *Let $m \geq n$, $c_1, \ldots, c_m > 0$ be such that $\sum_{i=1}^m c_i = n$ and $v_1, \ldots, v_m \in \mathbb{R}^n$. Let E and F be the best constants such that foll every integrable functions $f_1, \ldots, f_m : \mathbb{R} \to [0, \infty)$, we have*

$$I(f_1, \ldots, f_m) \geq E \cdot \prod_{i=1}^m \left(\int_{\mathbb{R}} f_i \right)^{c_i}, \tag{8.1}$$

$$J(f_1, \ldots, f_m) \leq F \cdot \prod_{i=1}^m \left(\int_{\mathbb{R}} f_i \right)^{c_i}. \tag{8.2}$$

Let E_g and F_g be the best constants such that these inequalities hold for all centred Gaussian functions of the form $f_i(t) = e^{-\alpha_i t^2}$ with any $\alpha_i > 0$. Let D be the best constant such that for every $\alpha_i > 0$, we have

$$\det\left(\sum_{i=1}^{m} \alpha_i c_i v_i v_i^T\right) \geq D \cdot \prod_{i=1}^{m} \alpha_i^{c_i}. \tag{8.3}$$

Then,

$$E = E_g = \sqrt{D} \qquad \text{and} \qquad F = F_g = \frac{1}{\sqrt{D}}. \tag{8.4}$$

8.2 Remark. There is a generalisation of this theorem concerning functions f_i defined on $\mathbb{R}^{n_i}$ for any $1 \leq n_i \leq n$ such that $n = \sum c_i n_i$ (the vectors v_i are replaced with linear maps $\mathbb{R}^n \to \mathbb{R}^{n_i}$).

8.3 Remark. As an example, consider the special case when $m = 2$, $n = 1$, $c_1 + c_2 = 1$ and $v_1 = v_2 = 1$. Then (8.3) becomes $\alpha_1 c_1 + \alpha_2 c_2 \geq D \cdot \alpha_1^{c_1} \alpha_2^{c_2}$ which holds with $D = 1$, which is sharp, by the AM-GM inequality. Thus $E = F = 1$ and (8.1) becomes the Prékopa-Leindler inequality (Theorem 2.8), whereas (8.2) becomes Hölder's inequality. Note that similarly, $D = 1$ when $n = 1$ with any m.

Theorem 8.1 will be established through several lemmas. Before we start, we remark that the theorem clearly holds when the v_i are linearly dependent because then $D = 0$. Moreover, $I = 0$ (regardless the f_i, the integrand will be a function defined on a lower dimensional subspace) and $J = \infty$. Thus we assume in what follows that the v_i are linearly dependent.

Our first lemma is a straighforward computation involving Gaussian functions. Recall that for $\alpha > 0$,

$$\int_{\mathbb{R}} e^{-\alpha t^2} \mathrm{d}t = \sqrt{\frac{\pi}{\alpha}}$$

and for a symmetric positive definite $n \times n$ matrix A,

$$\int_{\mathbb{R}^n} e^{-\langle Ax, x\rangle} \mathrm{d}x = \frac{\pi^{n/2}}{\sqrt{\det A}}.$$

8.4 Lemma. $F_g = \frac{1}{\sqrt{D}}$.

Proof. Let $f_i(t) = e^{-\alpha_i t^2}$, $\alpha_i > 0$, $i \leq m$. Then,

$$\prod_{i=1}^{m} \left(\int_{\mathbb{R}} f_i\right)^{c_i} = \prod_{i=1}^{m} \sqrt{\frac{\pi}{\alpha_i}}^{c_i} = \frac{\pi^{n/2}}{\sqrt{\prod \alpha_i^{c_i}}}$$

and

$$J(f_1, \ldots, f_m) = \int_{\mathbb{R}^n} \prod_{i=1}^{m} e^{-\alpha_i c_i \langle x, v_i\rangle^2} \mathrm{d}x = \int_{\mathbb{R}^n} e^{-\langle (\sum \alpha_i c_i v_i v_i^T) x, x\rangle} \mathrm{d}x = \frac{\pi^{n/2}}{\sqrt{\det\left(\alpha_i c_i v_i v_i^T\right)}}.$$

Therefore, F_g is the best constant in the inequality

$$\frac{\pi^{n/2}}{\sqrt{\det\left(\alpha_i c_i v_i v_i^T\right)}} \leq F_g \cdot \frac{\pi^{n/2}}{\sqrt{\prod \alpha_i^{c_i}}},$$

that is

$$\det\left(\alpha_i c_i v_i v_i^T\right) \geq \frac{1}{F_g^2} \cdot \prod_{i=1}^{m} \alpha_i^{c_i}.$$

Thus by the definition of D, $D = \frac{1}{F_g^2}$. □

8.5 Lemma. $E_g \cdot F_g = 1$.

Proof. For $\alpha_1, \ldots, \alpha_m > 0$ define a positive definite matrix

$$Q = \sum_{i=1}^{m} \alpha_i c_i v_i v_i^T$$

and the norm

$$N(x) = \sqrt{\langle Qx, x\rangle}$$

whose unit ball is the ellipsoid given by Q. Note that for $f_i(t) = e^{-\alpha_i t^2}$, as computed in Lemma 8.4, we have

$$F_g(\alpha_1, \ldots, \alpha_m) = \frac{J(f_1, \ldots, f_m)}{\prod_{i=1}^{m}\left(\int_{\mathbb{R}} f_i\right)^{c_i}} = \sqrt{\frac{\prod \alpha_i^{c_i}}{\det Q}}.$$

On the other hand,

$$E_g(\alpha_1, \ldots, \alpha_m) = \frac{I(f_1, \ldots, f_m)}{\prod_{i=1}^{m}\left(\int_{\mathbb{R}} f_i\right)^{c_i}} = \sqrt{\frac{\prod \alpha_i^{c_i}}{\pi^n}} \int_{\mathbb{R}^n} e^{-\inf\left\{\sum \alpha_i c_i t_i^2,\ x = \sum c_i t_i v_i\right\}} \mathrm{d}x.$$

Let us try to interpret the function given by the infimum that shows up in the exponent. Consider the dual norm,

$$N_\star(x) = \sup\left\{\langle x, y\rangle,\ N(y) \leq 1\right\}$$

and compute it explicitly. As the dual to $N(x) = \sqrt{\langle Qx, x\rangle}$ whose unit ball is an ellipsoid, $N_\star$ should also be of this form (duals of ellipsoids are ellipsoids). Note that by the Cauchy-Schwarz inequality,

$$\langle x, y\rangle = \left\langle Q^{-1/2}x, Q^{1/2}y\right\rangle \leq |Q^{-1/2}x| \cdot |Q^{1/2}y| = \sqrt{\langle Q^{-1}x, x\rangle}\sqrt{\langle Qy, y\rangle},$$

which gives, after taking the supremum over y such that $N(y) \leq 1$, that is $\langle Qy, y\rangle \leq 1$,

$$N_\star(x) \leq \sqrt{\langle Q^{-1}x, x\rangle}.$$

In fact, there is equality because we can arrange the Cauchy-Schwarz inequality to be equality by taking $y = \lambda Q^{-1}x$ (and then choosing λ such that $N(y) = 1$).

Let us now compute $N_\star$ differently. The condition $N(y) \leq 1$ reads $\sum \alpha_i c_i \langle y, v_i\rangle^2 \leq 1$. For x of the form $x = \sum c_i t_i v_i$, by the Cauchy-Schwarz inequality,

$$\langle x, y\rangle = \sum c_i t_i \langle y, v_i\rangle \leq \sqrt{\sum c_i \frac{t_i^2}{\alpha_i}}\sqrt{\sum \alpha_i c_i \langle y, v_i\rangle^2},$$

which shows that

$$N_\star(x) \le \inf\left\{\sqrt{\sum c_i \frac{t_i^2}{\alpha_i}},\ x = \sum c_i t_i v_i\right\}.$$

In fact, this is equality. To see that, given x, choose y achieving the supremum in the definition of $N_\star$, that is $y = \lambda Q^{-1}x$ with λ such that $N(y) = 1$. Let $t_i = \frac{1}{\lambda}\alpha_i\langle y, v_i\rangle$. Then we have equality in the Cauchy-Schwarz inequality above. Moreover, for these t_i we have $x = \frac{1}{\lambda}Qy = \frac{1}{\lambda}\sum \alpha_i c_i \langle y, v_i\rangle v_i = \sum t_i c_i v_i$, which finishes the argument.

Going back the the formula for $E_g(\alpha_1, \ldots, \alpha_m)$, we can rewrite it as

$$E_g(\alpha_1^{-1}, \ldots, \alpha_m^{-1}) = \sqrt{\frac{\prod \alpha_i^{-c_i}}{\pi^n}} \int_{\mathbb{R}^n} e^{-N_\star(x)^2} \mathrm{d}x.$$

Since $N_\star(x) = \sqrt{\langle Q^{-1}x, x\rangle}$, we get

$$E_g(\alpha_1^{-1}, \ldots, \alpha_m^{-1}) = \sqrt{\frac{\prod \alpha_i^{-c_i}}{\pi^n}} \int_{\mathbb{R}^n} e^{-\langle Q^{-1}x, x\rangle} \mathrm{d}x = \sqrt{\frac{\prod \alpha_i^{-c_i}}{\det Q^{-1}}}.$$

We thus obtain

$$F_g(\alpha_1, \ldots, \alpha_m) \cdot E_g(\alpha_1^{-1}, \ldots, \alpha_m^{-1}) = \sqrt{\frac{\prod \alpha_i^{c_i}}{\det Q}} \cdot \sqrt{\frac{\prod \alpha_i^{-c_i}}{\det Q^{-1}}} = 1,$$

that is

$$F_g(\alpha_1, \ldots, \alpha_m) = \frac{1}{E_g(\alpha_1^{-1}, \ldots, \alpha_m^{-1})}.$$

Taking the supremum over $\alpha_i > 0$ gives $F_g = \frac{1}{E_g}$. □

8.6 Lemma. *For every integrable functions $f_1, \ldots, f_m, h_1, \ldots, h_m : \mathbb{R} \to [0, \infty)$ with $\int f_i = 1 = \int h_i$, we have*

$$I(f_1, \ldots, f_m) \ge D \cdot J(h_1, \ldots, h_m).$$

Note that this lemma gives $E \ge DF$, so by Lemma 8.4 and 8.5,

$$\sqrt{D} = E_g \ge E \ge DF \ge DF_g = D\frac{1}{\sqrt{D}} = \sqrt{D},$$

so there are in fact equalities and this finishes the proof of Theorem 8.1.

Proof of Lemma 8.6. If $D = 0$, there is nothing to prove, so let D be positive. Without loss of generality we can assume that the f_i and h_i are positive and continuous (...). Define the transport functions $T_i : \mathbb{R} \to \mathbb{R}$ by

$$\int_{-\infty}^{T_i(t)} f_i = \int_{-\infty}^{t} h_i.$$

Then, differentiating yields the transport equations,

$$T_i'(t) f_i(T_i(t)) = h_i(t).$$

Define the change of variables $\Psi : \mathbb{R}^n \to \mathbb{R}^n$,

$$\Psi(y) = \sum_{i=1}^{m} c_i T_i(\langle\langle y, v_i\rangle\rangle) v_i.$$

Since

$$\frac{\partial \Psi}{\partial y_j} = \sum_i c_i T_i'(\langle\langle y, v_i\rangle\rangle) v_{i,j} v_i,$$

we have

$$d\Psi(y) = \sum_i c_i T_i'(\langle\langle y, v_i\rangle\rangle) v_i v_i^T$$

and because T_i' is pointwise positive, $d\Psi$ is positive definite and thus has a positive determinant. Therefore Ψ is injective, $x = \Psi(y)$ defines a valid change of variables on $\mathbb{R}^n$ and we have

$$I(f_1, \ldots, f_m) = \int_{\mathbb{R}^n} \sup_{x=\sum c_i t_i v_i} \prod f_i(t_i)^{c_i} \mathrm{d}x \geq \int_{\mathbb{R}^n} \sup_{\Psi(y)=\sum c_i t_i v_i} \prod f_i(t_i)^{c_i} \det(d\Psi(y)) \mathrm{d}y.$$

By (8.3), $\det(d\Psi(y)) \geq D \cdot \prod T_i'(\langle\langle y, v_i\rangle\rangle)^{c_i}$. Setting in the supremum $t_i = T_i(\langle\langle y, v_i\rangle\rangle)$ and using the transport equations we obtain

$$\begin{aligned} I(f_1, \ldots, f_m) \geq D \int_{\mathbb{R}^n} \prod f_i(T_i(\langle\langle y, v_i\rangle\rangle))^{c_i} \prod T_i'(\langle\langle y, v_i\rangle\rangle)^{c_i} \mathrm{d}y &= D \int_{\mathbb{R}^n} \prod h_i(\langle\langle y, v_i\rangle\rangle)^{c_i} \mathrm{d}y \\ &= D \cdot J(h_1, \ldots, h_m). \end{aligned}$$

□

8.2 Geometric applications

Suppose we are given positive numbers $c_1, \ldots, c_m > 0$ and unit vectors $v_1, \ldots, v_m$ in $\mathbb{R}^n$ such that $\sum c_i v_i v_i^T = I$. Recall Theorem 5.14 says that then

$$\int_{\mathbb{R}^n} \prod f_i(\langle\langle x, v_i\rangle\rangle)^{c_i} \mathrm{d}x \leq \prod \left(\int_{\mathbb{R}} f_i \right)^{c_i}. \tag{8.5}$$

Recall also that automatically, $\sum c_i = n$ (Remark 5.5). In terms of Theorem 8.1, this means $F \leq 1$. We shall prove now that $D = 1$, thus $F = 1$, thus giving also a proof of Theorem 5.14.

8.7 Theorem. *If positive numbers $c_1, \ldots, c_m > 0$ and unit vectors $v_1, \ldots, v_m$ in $\mathbb{R}^n$ are such that $\sum_{i=1}^m c_i v_i v_i^T = I$, then for every positive $\alpha_1, \ldots, \alpha_m$,*

$$\det\left(\sum_{i=1}^{m} \alpha_i c_i v_i v_i^T \right) \geq \prod_{i=1}^{m} \alpha_i^{c_i}$$

(which is sharp for $\alpha_i = 1$, $i \leq m$).

We shall need two basic tools from multilinear algebra.

8.8 Theorem (Cauchy-Binet formula)**.** *Let $n \le m$. Let A and B be $n \times m$ and $m \times n$ matrices. Then*

$$\det(AB) = \sum_{|S|=n} \det(A_S)\det(B^S),$$

where the sum is over all n-element subsets S of the set $\{1, \dots, m\}$ and A_S, B^S denotes the restriction of A, respectively B to an $n \times n$ matrix with colums, respectively rows from S.

8.9 Theorem (Sylvester's formula)**.** *Let A and B be $n \times m$ and $m \times n$ matrices. Then*

$$\det(I_n + AB) = \det(I_m + BA).$$

Proof of Theorem 8.7. Let V be an $n \times m$ matrix with columns $\sqrt{\alpha_i c_i} v_i$, $i \le m$. Then

$$VV^T = \sum_{i=1}^m \alpha_i c_i v_i v_i^T,$$

so by the Cauchy-Binet formula,

$$\det\left(\sum_{i=1}^m \alpha_i c_i v_i v_i^T\right) = \det(VV^T) = \sum_{|S|=n} (\det V_S)^2.$$

Note that $\det V_S = \left(\prod_{i\in S}\sqrt{\alpha_i}\right)\det\left([\sqrt{c_i}v_i]_{i\in S}\right)$, where $[\sqrt{c_i}v_i]_{i\in S}$ is the $n \times n$ matrix with columns $\sqrt{c_i}u_i$, $i \in S$. Denoting

$$\alpha_S = \prod_{i\in S}\alpha_i$$

and

$$\lambda_S = \det\left([\sqrt{c_i}v_i]_{i\in S}\right)^2,$$

we thus have

$$\det\left(\sum_{i=1}^m \alpha_i c_i v_i v_i^T\right) = \sum_{|S|=n}\lambda_S\alpha_S.$$

When the α_i are all 1, the above becomes $1 = \sum_{|S|=n}\lambda_S$. Hence, by the AM-GM inequality,

$$\sum_{|S|=n}\lambda_S\alpha_S \ge \prod_{|S|=n} a_S^{\lambda_S} = \prod_{|S|=n}\prod_{i\in S}\alpha_i^{\lambda_S} = \prod_{|S|=n}\prod_{i=1}^m \alpha_i^{\lambda_S \mathbf{1}_{i\in S}} = \prod_{i=1}^m a_i^{\sum_{S:i\in S}\lambda_S}.$$

To finish, note that for a fixed i,

$$\sum_{S:i\in S}\lambda_S = \sum_S \lambda_S - \sum_{S:i\notin S}\lambda_S = 1 - \sum_{S:i\notin S}\left(\det\left([\sqrt{c_j}v_j]_{j\in S}\right)\right)^2.$$

Using the Cauchy-Binet formula once again,

$$\begin{aligned}\sum_{S:i\notin S}\left(\det\left([\sqrt{c_j}v_j]_{j\in S}\right)\right)^2 = \det\left(\sum_{j\ne i} c_j v_j v_j^T\right) &= \det\left(I_n - c_i v_i v_i^T\right) \\ &= 1 - c_i v_i^T v_i = 1 - c_i|v_i|^T = 1 - c_i,\end{aligned}$$

where in the second line we use Sylvester's formula. Thus,

$$\sum_{S:i\in S} \lambda_S = c_i$$

and consequently,

$$\det\left(\sum_{i=1}^m \alpha_i c_i v_i v_i^T\right) = \sum_{|S|=n} \lambda_S \alpha_S \geq \prod_{|S|=n} a_S^{\lambda_S} \prod_{i=1}^m a_i^{\sum_{S:i\in S}\lambda_S} = \prod_{i=1}^m a_i^{c_i}.$$

□

We saw how useful Ball's version (8.5) of the Brascamp-Lieb inequality (8.2) was to solve the reverse isoperimetric problem. We finish with one more application, which can be viewed as a sharp version for Gaussian vectors of the crucial Lemma 6.5 (used for the proof of Dvoretzky's theorem).

8.10 Theorem. *Let G be a standard Gaussian vector in $\mathbb{R}^n$. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ whose unit ball is in John's position. Then $\mathbb{E}\|G\| \geq \mathbb{E}\|G\|_\infty$.*

Proof. Let $c_1, \ldots, c_m > 0$ and contact points $v_1, \ldots, v_m$ be given by John's theorem 5.3, $I = \sum c_i v_i v_i^T$. As in the proof of Ball's theorem 5.6, the unit ball of $\|\cdot\|$ is contained in

$$\{x \in \mathbb{R}^n,\ |\langle x, v_i\rangle| \leq 1, i \leq m\}$$

which is the unit ball with respect to

$$\|x\|' = \max_{i\leq m} |\langle x, v_i\rangle|.$$

Thus, $\|x\| \geq \|x\|'$ and we get

$$\begin{aligned}\mathbb{E}\|G\| \geq \mathbb{E}\|G\|' = \mathbb{E}\max_{i\leq m}|\langle G, v_i\rangle| &= \int_0^\infty \mathbb{P}\left(\max_{i\leq m}|\langle G, v_i\rangle| > s\right)\mathrm{d}s \\ &= \int_0^\infty \left[1 - \mathbb{P}\left(\forall i \leq m\ |\langle G, v_i\rangle| \leq s\right)\right]\mathrm{d}s\end{aligned}$$

Applying (8.5) to $f_i(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\mathbf{1}_{[-s,s]}(t)$ gives (recall (5.4))

$$\begin{aligned}\mathbb{P}\left(\forall i \leq m\ |\langle G, v_i\rangle| \leq t\right) &= \int_{\mathbb{R}^n} \frac{1}{\sqrt{2\pi}^n} e^{-|x|^2/2} \prod_{i=1}^m \mathbf{1}_{\{|\langle x, v_i\rangle| \leq s\}}\mathrm{d}x \\ &= \int_{\mathbb{R}^n} \prod_{i=1}^m f_i(\langle x, v_i\rangle)^{c_i}\mathrm{d}x \\ &\leq \prod_{i=1}^m \left(\int_{\mathbb{R}} f_i\right)^{c_i} \\ &= \prod_{i=1}^m \mathbb{P}\left(|G_1| \leq s\right)^{c_i} = \mathbb{P}\left(|G_1| \leq s\right)^n.\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}\|G\| &\geq \int_0^\infty \left[1 - \mathbb{P}\left(\forall i \leq m \ |\langle G, v_i \rangle| \leq s\right)\right] \mathrm{d}s \\
&\geq \int_0^\infty \left[1 - \mathbb{P}\left(|G_1| \leq s\right)^n\right] \mathrm{d}s \\
&= \int_0^\infty \left[1 - \mathbb{P}\left(\forall i \leq n \ |G_i| \leq s\right)\right] \mathrm{d}s \\
&= \int_0^\infty \mathbb{P}\left(\max_{i \leq n} \ |G_i| > s\right) \mathrm{d}s \\
&= \mathbb{E}\|G\|_\infty.
\end{aligned}$$

□

8.3 Applications in analysis: Young's inequalities

We shall derive classical Young's iequalities for L_p norms of convolutions with sharp constants from the Brascamp-Lieb inequality (8.2). Recall that for a measurable function $f : \mathbb{R} \to \mathbb{R}$ and $p \in [1, \infty]$, its L_p norm is defined as

$$\|f\|_p = \left(\int_{\mathbb{R}} |f|^p\right)^{1/p}.$$

By $p' = \frac{p}{p-1}$ we denote the dual exponent, $\frac{1}{p} + \frac{1}{p'} = 1$. It follows from Hölder's inequality that we have the following variational formula

$$\|f\|_p = \sup\left\{\int_{\mathbb{R}} fh, \ \int_{\mathbb{R}} |h|^{p'} \leq 1\right\}. \tag{8.6}$$

8.11 Theorem (Young's inequality)**.** *Let $p, q, r \geq 1$ be such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$. Then for every measurable functions $f, g : \mathbb{R} \to \mathbb{R}$, we have*

$$\|f \star g\|_r \leq \frac{C_p C_q}{C_r} \|f\|_p \cdot \|g\|_q, \tag{8.7}$$

where

$$C_p = \sqrt{\frac{p^{1/p}}{p'^{1/p'}}}. \tag{8.8}$$

Moreover, this inequality is sharp.

8.12 Remark. Clearly, $C_{p'} = \frac{1}{C_p}$.

8.13 Remark. Using the variational formula, (8.7) can be equivalently stated as: for every $p, q, r' \geq 1$ such that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r'} = 2$ and $f, g, h : \mathbb{R} \to \mathbb{R}$, we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y) g(y) h(x) \mathrm{d}x \mathrm{d}y \leq C_p C_q C_r \|f\|_p \|g\|_q \|h\|_{r'}. \tag{8.9}$$

To see this, note that by (8.6),

$$\|f \star g\|_r = \sup_h \frac{\int_{\mathbb{R}} (f \star g)(x) h(x) \mathrm{d}x}{\|h\|_{r'}} = \sup_h \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} f(y-x) g(y) h(x) \mathrm{d}x \mathrm{d}y}{\|h\|_{r'}}.$$

Combining this with Remark 8.12 finishes the argument.

8.14 Remark. Setting $\tilde{h}(x) = h(-x)$, the left hand side of (8.9) can be written as

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y)g(y)h(x)\mathrm{d}x\mathrm{d}y = (f \star g \star \tilde{h})(0).$$

Thus, (8.9), equivalently (8.7) follow from the following: for every $p, q, r' \geq 1$ such that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r'} = 2$ and $f, g, h : \mathbb{R} \to \mathbb{R}$, we have

$$\|f \star g \star h\|_\infty \leq C_p C_q C_r \|f\|_p \|g\|_q \|h\|_{r'}. \tag{8.10}$$

In fact, the three are equivalent: from Hölder's inequality and (8.7), we get

$$\|f \star g \star h\|_\infty \leq \|f \star g\|_r \|h\|_{r'} \leq \frac{C_p C_q}{C_r} \|f\|_p \|g\|_q \|h\|_{r'} = C_p C_q C_{r'} \|f\|_p \|g\|_q \|h\|_{r'}.$$

The gain is that (8.10) easily tensorises, say we have $f, g, h : \mathbb{R}^2 \to \mathbb{R}$. Then using it twice, for every $(x_1, x_2) \in \mathbb{R}^2$ we obtain

$$\begin{aligned}
(f \star g \star h)(x_1, x_2) &= \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(x_1 - y_1 - z_1, x_2 - y_2 - z_2) g(y_1, y_2) h(z_1, z_2) \mathrm{d}y \mathrm{d}z \\
&\leq \int_{\mathbb{R}^2} (C_p C_q C_{r'}) \|f(\cdot, x_2 - y_2 - z_2)\|_p \|g(\cdot, y_2)\|_q \|h(\cdot, z_2)\|_{r'} \mathrm{d}y_2 \mathrm{d}z_2 \\
&\leq (C_p C_q C_{r'})^2 \|f\|_p \|g\|_q \|h\|_{r'}.
\end{aligned}$$

Therefore, once we have proved (8.9) for functions on $\mathbb{R}$, then by iterating the argument above, we get that (8.10) holds for all functions on $\mathbb{R}^n$ with the constant $(C_p C_q C_{r'})^n$ and by the established equivalences, (8.7) and (8.9) also hold for all functions on $\mathbb{R}^n$ (with constants being the nth-power).

Proof of (8.9). Replacing f, g, h with $f^{1/p}$, $g^{1/q}$ and $h^{1/r'}$, introducing $c_1 = 1/p$, $c_2 = 1/q$, $c_3 = 1/r'$ and the vectors $v_1 = (1, -1)$, $v_2 = (0, 1)$, $v_3 = (1, 0)$, note that (8.9) becomes

$$\int_{\mathbb{R}^2} f(\langle\langle x, v_1 \rangle\rangle)^{c_1} g(\langle\langle x, v_2 \rangle\rangle)^{c_2} h(\langle\langle x, v_3 \rangle\rangle)^{c_3} \mathrm{d}x \leq F \cdot \left(\int f\right)^{c_1} \left(\int g\right)^{c_2} \left(\int h\right)^{c_3}$$

with $c_1 + c_2 + c_3 = 2$. This is exactly the Brascamp-Lieb framework and by Theorem 8.1, the best constant F is $1/\sqrt{D}$ with D being the best constant in the inequality

$$\det(\alpha_1 c_1 v_1 v_1^T + \alpha_2 c_2 v_2 v_2^T + \alpha_3 c_3 v_3 v_3^T) \geq D \cdot \alpha_1^{c_1} \alpha_2^{c_2} \alpha_3^{c_3},$$

for every $\alpha_1, \alpha_2, \alpha_3 > 0$. We have

$$\begin{aligned}
\det(\alpha_1 c_1 v_1 v_1^T + \alpha_2 c_2 v_2 v_2^T + \alpha_3 c_3 v_3 v_3^T) &= \det\left(\alpha_1 c_1 \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right] + \alpha_2 c_2 \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right] + \alpha_3 c_3 \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]\right) \\
&= \det(\left[\begin{smallmatrix} \alpha_1 c_1 + \alpha_3 c_3 & -\alpha_1 c_1 \\ -\alpha_1 c_1 & \alpha_1 c_1 + \alpha_2 c_2 \end{smallmatrix}\right]) \\
&= \alpha_1 \alpha_2 c_1 c_2 + \alpha_2 \alpha_3 c_2 c_3 + \alpha_1 \alpha_3 c_1 c_3.
\end{aligned}$$

By taking the limit if necessary, we can assume that $c_1, c_2, c_3 < 1$. Then by the AM-GM inequality,

$$\begin{aligned}\alpha_1\alpha_2 c_1 c_2 + \alpha_2\alpha_3 c_2 c_3 + \alpha_1\alpha_3 c_1 c_3 &= (1-c_3)\frac{c_1c_2}{1-c_3}\alpha_1\alpha_2 \\ &+ (1-c_1)\frac{c_2c_3}{1-c_1}\alpha_2\alpha_3 \\ &+ (1-c_2)\frac{c_1c_3}{1-c_2}\alpha_1\alpha_3 \\ &\geq \left(\frac{c_1c_2}{1-c_3}\right)^{1-c_3}\left(\frac{c_2c_3}{1-c_1}\right)^{1-c_1}\left(\frac{c_1c_3}{1-c_2}\right)^{1-c_2} \\ &\quad \cdot \alpha_1^{2-c_2-c_3}\alpha_2^{2-c_1-c_3}\alpha_3^{2-c_1-c_2} \\ &= \prod_{i=1}^3 \frac{c_i^{c_i}}{(1-c_i)^{1-c_i}} \cdot \alpha_1^{c_1}\alpha_2^{c_2}\alpha_3^{c_3}.\end{aligned}$$

Equality holds for α_i such that

$$\frac{c_1c_2}{1-c_3}\alpha_1\alpha_2 = \frac{c_2c_3}{1-c_1}\alpha_2\alpha_3 = \frac{c_1c_3}{1-c_2}\alpha_1\alpha_3.$$

Therefore the best constant is

$$F = \frac{1}{\sqrt{D}} = \prod_{i=1}^3 \sqrt{\frac{(1-c_i)^{1-c_i}}{c_i^{c_i}}}$$

which is $C_pC_qC_{r'}$, as required (recall $c_1 = 1/p$, so $1 - c_1 = 1/p'$, etc.). $\square$

8.4 Applications in information theory: entropy power

The (differential, Shannon) **entropy** of a random vector X in $\mathbb{R}^n$ with density $f : \mathbb{R}^n \to [0, +\infty)$ is defined by

$$\mathcal{S}(X) = -\int_{\mathbb{R}^n} f \log f$$

(provided that the integral exists in the usual Lebesgue's sense). For $p > 0$, $p \neq 1$, we also define the **p-Rényi entropy** of X as

$$\mathcal{S}_p(X) = \frac{1}{1-p}\log\int_{\mathbb{R}^n} f^p.$$

This is a well defined quantity (the Lebesgue integral of f^p always exists). However, it can be that $\mathcal{S}_p$ is infinite for every $p \neq 1$, e.g. take f on $\mathbb{R}$ to be proportional to $\frac{1}{x\ln^2 x}\mathbf{1}_{(0,1/2)\cup(2,\infty)}$. Note that for this density, the entropy $\mathcal{S}(X)$ does not exist.

Remark that

$$\mathcal{S}(X) = \mathbb{E}\big[-\log f(X)\big].$$

Similarly,

$$\mathcal{S}_p(X) = \frac{1}{1-p}\log\mathbb{E}f(X)^{p-1} = -\log(\mathbb{E}f(X)^{p-1})^{\frac{1}{p-1}},$$

so in particular, $\mathcal{S}_p(X)$ is nonincreasing in p. By monotonicity, the one sided limits $\lim_{p\to 1\pm}\mathcal{S}_p(X)$ do exist, but possibly taking different values (as in the mentioned example, they are $\mp\infty$).

If the Rényi entropy $\mathcal{S}_p(X)$ is finite at some point $p = p_0 > 1$, then $\mathcal{S}(X)$ exits and

$$\lim_{p\to 1+} \mathcal{S}_p(X) = \mathcal{S}(X).$$

Similarly, if $\mathcal{S}_p(X)$ is finite for some $p = p_0 < 1$, then $\mathcal{S}(X)$ exists and equals the limit $\lim_{p\to 1-} \mathcal{S}_p(X)$.

For example, for X with a density proportional to $\frac{1}{x\ln^3 x}\mathbf{1}_{(2,\infty)}$, its Rényi entropy of order p is $+\infty$ when $p \in (0,1)$ and is finite when $p \in (1,\infty)$, its entropy $\mathcal{S}(X)$ exists, is finite and equals $\lim_{p\to 1+} \mathcal{S}_p(X)$, but $\lim_{p\to 1-} \mathcal{S}_p(X) = +\infty$.

We finish this discussion of definiteness with one more remark. If a density f belongs to $L_p(\mathbb{R}^n)$ for some $p > 1$, then by the concavity of the logarithm and Jensen's inequality $-\int_{\mathbb{R}^n} f\log f > -\infty$. Moreover, Jensen's inequality also yields the following very useful variational formula for the entropy

$$\mathcal{S}(X) = \inf\left\{-\int_{\mathbb{R}^n} f\log g,\ g : \mathbb{R}^n \to [0,+\infty) \text{ is a probability density}\right\}.$$

In particular, if $\mathbb{E}|X|^2 < \infty$, then comparison with the standard Gaussian gives $\mathcal{S}(X) < \infty$. Combining the two we can conclude that the entropy of a log-concave random vector with density on $\mathbb{R}^n$ is well defined and finite.

We remark the following scaling properties: for a linear invertible map $A : \mathbb{R}^n \to \mathbb{R}^n$ and $b \in \mathbb{R}^n$, we have

$$\mathcal{S}(AX + b) = \mathcal{S}(X) + \log|\det A|$$

and identically,

$$\mathcal{S}_p(AX + b) = \mathcal{S}_p(X) + \log|\det A|.$$

For instance, for a standard Gaussian vector G in $\mathbb{R}^n$, we have

$$\begin{aligned}\mathcal{S}(G) &= -\int_{\mathbb{R}^n} \frac{1}{\sqrt{2\pi}^n} e^{-|x|^2/2} \log\left(\frac{1}{\sqrt{2\pi}^n} e^{-|x|^2/2}\right) \mathrm{d}x \\ &= \log\sqrt{2\pi}^n + \frac{n}{2} = \frac{n}{2}\log(2\pi e).\end{aligned}$$

Consequently, for a Gaussian vector G_A with covariance matrix A, $G_A = A^{1/2}G$,

$$\mathcal{S}(G_A) = \mathcal{S}(G) + \log|\det A^{1/2}| = \frac{n}{2}\log\left[2\pi e(\det A)^{1/n}\right]. \tag{8.11}$$

The fundamental inequality in information theory, put forward by Shannon, the so-called entropy power inequality, concerns a subadditivity property of the entropy of sums of independent random vectors vectors.

8.15 Theorem (Entropy power inequality)**.** *Let X and Y be independent random vectors in $\mathbb{R}^n$, let $\lambda \in [0,1]$. Then (provided that all the entropies exist)*

$$\mathcal{S}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda\mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y). \tag{8.12}$$

Proof. Let f and g be the densities of X and Y. For $a > 0$, let f_a denote the density of aX. The density of $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$ is $f_{\sqrt{\lambda}} \star g_{\sqrt{1-\lambda}}$. By Young's inequality (8.7) (in $\mathbb{R}^n$, see Remark 8.14),

$$\|f_{\sqrt{\lambda}} \star g_{\sqrt{1-\lambda}}\|_r \leq \left(\frac{C_p C_q}{C_r}\right)^n \|f_{\sqrt{\lambda}}\|_p \|g_{\sqrt{1-\lambda}}\|_q,$$

for all $p, q, r > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Taking the log and rewriting in terms of Rényi entropy yields

$$\begin{aligned}\mathcal{S}_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq & \frac{r}{1-r} n \log \frac{C_p C_q}{C_r} \\ & + \frac{r}{1-r}\frac{1-p}{p}\mathcal{S}_p(\sqrt{\lambda}X) + \frac{r}{1-r}\frac{1-q}{q}\mathcal{S}_q(\sqrt{1-\lambda}Y).\end{aligned}$$

We have $\mathcal{S}_p(\sqrt{\lambda}X) = \frac{n}{2}\log\lambda + \mathcal{S}_p(X)$ and similarly for $\sqrt{1-\lambda}Y$. Given $r > 1$, take $p, q > 1$ such that

$$\frac{1}{p} = \frac{\lambda}{r} + 1 - \lambda \qquad \text{and} \qquad \frac{1}{q} = \frac{1-\lambda}{r} + \lambda.$$

Then the condition $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ is satisfied. Note that

$$\frac{r}{1-r}\frac{1-p}{p} = \lambda \qquad \text{and} \qquad \frac{r}{1-r}\frac{1-q}{q} = 1 - \lambda.$$

Putting these together gives

$$\begin{aligned}\mathcal{S}_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda\mathcal{S}_p(X) - (1-\lambda)\mathcal{S}_q(Y) \geq & \frac{n}{2}\Big[\frac{r}{1-r}\log\left(\frac{C_p C_q}{C_r}\right)^2 \\ & + \lambda\log\lambda + (1-\lambda)\log(1-\lambda)\Big].\end{aligned}$$

As $r \to 1$, we also have $p, q \to 1$, and consequently the left hand side converges to $\mathcal{S}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda\mathcal{S}(X) - (1-\lambda)\mathcal{S}(Y)$. It is enough to show that the right hand side converges to 0. We have

$$\frac{r}{1-r}\log\left(\frac{C_p C_q}{C_r}\right)^2 = -r'\log\left(\frac{p^{1/p}}{p'^{1/p'}}\frac{q^{1/q}}{q'^{1/q'}}\frac{r^{-1/r}}{r'^{-1/r'}}\right).$$

Using that $\frac{r'}{p'} = \lambda$ and $\frac{r'}{q'} = 1 - \lambda$, after a few simplifications, we get to

$$\begin{aligned}-r'\log\left(\frac{p^{1/p}}{p'^{1/p'}}\frac{q^{1/q}}{q'^{1/q'}}\frac{r^{1/r}}{r'^{1/r'}}\right) = & r'\left(\frac{1}{p}\log\frac{1}{p} + \frac{1}{q}\log\frac{1}{q} - \frac{1}{r}\log\frac{1}{r}\right) \\ & - \lambda\log\lambda - (1-\lambda)\log(1-\lambda).\end{aligned}$$

The term $\lambda\log\lambda + (1-\lambda)\log(1-\lambda)$ cancels and we are left with showing that the first term goes to 0. Setting $r' = \frac{1}{\varepsilon}$, $\varepsilon \to 0$ and using $\frac{1}{p} = \frac{\lambda}{r} + 1 - \lambda = \lambda(1-\varepsilon) + 1 - \lambda = 1 - \lambda\varepsilon$, similarly $\frac{1}{q} = 1 - (1-\lambda)\varepsilon$, we obtain

$$\begin{aligned} r'\left(\frac{1}{p}\log\frac{1}{p} + \frac{1}{q}\log\frac{1}{q} - \frac{1}{r}\log\frac{1}{r}\right) &= \frac{1}{\varepsilon}\Big[(1-\lambda\varepsilon)\log(1-\lambda\varepsilon) \\ &\quad + (1-(1-\lambda)\varepsilon)\log(1-(1-\lambda)\varepsilon) \\ &\quad - (1-\varepsilon)\log(1-\varepsilon)\Big] \\ &\approx \frac{1}{\varepsilon}\Big[-(1-\lambda\varepsilon)\lambda\varepsilon \\ &\quad - (1-(1-\lambda)\varepsilon)(1-\lambda)\varepsilon \\ &\quad + (1-\varepsilon)\varepsilon\Big] \\ &= \Big[\lambda^2 + (1-\lambda)^2 - 1\Big]\varepsilon. \end{aligned}$$

□

The entropy power inequality is sometimes stated in another equivalent forms.

8.16 Theorem (Entropy power inequalities)**.** *The following statements holding true for every independent random vectors X and Y in $\mathbb{R}^n$ are equivalent*

(i) $\mathcal{S}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda\mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y)$, *for every* $\lambda \in [0,1]$,

(ii) $e^{\frac{2}{n}\mathcal{S}(X+Y)} \geq e^{\frac{2}{n}\mathcal{S}(X)} + e^{\frac{2}{n}\mathcal{S}(Y)}$,

(iii) $\mathcal{S}(X+Y) \geq \mathcal{S}(G_X+G_Y)$, *for independent Gaussian vectors G_X and G_Y in $\mathbb{R}^n$ with covariance matrices proportional to the identity matrix such that $\mathcal{S}(G_X) = \mathcal{S}(X)$ and $\mathcal{S}(G_Y) = \mathcal{S}(Y)$.*

Proof. "(i)⟹(ii)" Applying (i) to $X/\sqrt{\lambda}$ and $Y/\sqrt{1-\lambda}$, we get

$$\begin{aligned} \mathcal{S}(X+Y) = \mathcal{S}\left(\sqrt{\lambda}\frac{X}{\sqrt{\lambda}} + \sqrt{1-\lambda}\frac{Y}{\sqrt{1-\lambda}}\right) &\geq \lambda\mathcal{S}\left(\frac{X}{\sqrt{\lambda}}\right) + (1-\lambda)\mathcal{S}\left(\frac{Y}{\sqrt{1-\lambda}}\right) \\ &= \lambda\mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y) - \frac{n}{2}\Big[\lambda\log\lambda + (1-\lambda)\log(1-\lambda)\Big]. \end{aligned}$$

We optimise the right hand side over $\lambda \in (0,1)$. Computing the derivative and equating it to 0 gives

$$\mathcal{S}(X) - \mathcal{S}(Y) - \frac{n}{2}\Big[\log\lambda - \log(1-\lambda)\Big] = 0.$$

Solving for λ yields

$$\lambda = \frac{e^{\frac{2}{n}\mathcal{S}(X)}}{e^{\frac{2}{n}\mathcal{S}(X)} + e^{\frac{2}{n}\mathcal{S}(Y)}}.$$

Note that also

$$\mathcal{S}(X) - \frac{n}{2}\log\lambda = \mathcal{S}(Y) - \frac{n}{2}\log(1-\lambda).$$

Then

$$
\begin{aligned}
\mathcal{S}(X+Y) &\geq \lambda \mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y) - \frac{n}{2}\Big[\lambda \log \lambda + (1-\lambda)\log(1-\lambda)\Big] \\
&= \lambda\Big(\mathcal{S}(X) - \frac{n}{2}\log \lambda\Big) + (1-\lambda)\Big(\mathcal{S}(Y) - \frac{n}{2}\log(1-\lambda)\Big) \\
&= \mathcal{S}(X) - \frac{n}{2}\log \lambda \\
&= \frac{n}{2}\log\left(e^{\frac{2}{n}\mathcal{S}(X)} + e^{\frac{2}{n}\mathcal{S}(Y)}\right),
\end{aligned}
$$

which gives (ii).

"(ii)$\Longrightarrow$(iii)" Let G_X and G_Y have covariance matrices αI and βI. Then $G_X + G_Y$ has covariance matrix $(\alpha+\beta)I$. By (8.11) and (ii), we obtain

$$
\begin{aligned}
e^{\frac{2}{n}\mathcal{S}(X+Y)} \geq e^{\frac{2}{n}\mathcal{S}(X)} + e^{\frac{2}{n}\mathcal{S}(Y)} = e^{\frac{2}{n}\mathcal{S}(G_X)} + e^{\frac{2}{n}\mathcal{S}(G_Y)} &= 2\pi e\Big(\alpha+\beta\Big) \\
&= e^{\frac{2}{n}\mathcal{S}(G_X+G_Y)}.
\end{aligned}
$$

"(iii)$\Longrightarrow$(i)" Let G_X and G_Y be Gaussian vectors with covariance matrices αI and βI such that $\mathcal{S}(G_X) = \mathcal{S}(X)$ and $\mathcal{S}(G_Y) = \mathcal{S}(Y)$. Let $\lambda \in [0,1]$. Then $\mathcal{S}(\sqrt{\lambda}G_X) = \mathcal{S}(\sqrt{\lambda}X)$ and $\mathcal{S}(\sqrt{1-\lambda}G_Y) = \mathcal{S}(\sqrt{1-\lambda}Y)$, so by (iii) and the AM-GM, we obtain

$$
\begin{aligned}
e^{\frac{2}{n}\mathcal{S}(\sqrt{\lambda}X+\sqrt{1-\lambda}Y)} \geq e^{\frac{2}{n}\mathcal{S}(\sqrt{\lambda}G_X+\sqrt{1-\lambda}G_Y)} &= e^{\frac{2}{n}\mathcal{S}\left(N(0,(\lambda\alpha+(1-\lambda)\beta)I)\right)} \\
&= 2\pi e(\lambda\alpha + (1-\lambda)\beta)) \\
&\geq 2\pi e(\alpha^{\lambda}\beta^{1-\lambda}) \\
&= e^{\frac{2}{n}\lambda\mathcal{S}(G_X)}e^{\frac{2}{n}(1-\lambda)\mathcal{S}(G_Y)} \\
&= e^{\frac{2}{n}\left(\lambda\mathcal{S}(X)+(1-\lambda)\mathcal{S}(G_Y)\right)},
\end{aligned}
$$

which shows (i). □

8.5 Entropy and slicing

Let X be a log-concave random vector in $\mathbb{R}^n$ with density f. In the proof of Theorem 2.20, we showed that

$$
\log(e^{-n}\|f\|_\infty) \leq \int_{\mathbb{R}^n} f \log f \leq \log \|f\|_\infty.
$$

If X is isotropic, its isotropic constant equals $L_X = \|f\|_\infty^{1/n}$. Therefore, the above establishes that

$$
\log(e^{-1}L_X) \leq -\frac{1}{n}\mathcal{S}(X) \leq \log L_X.
$$

Showing that L_X is bounded above by a universal constant (the slicing problem – Conjecture 4.15) is thus equivalent to showing that $-\frac{1}{n}\mathcal{S}(X)$ is bounded above by a universal constant, assuming X is isotropic.

To make an affine invariant statement, let us introduce the **relative entropy** $\mathcal{D}(X)$ of X defined as

$$\mathcal{D}(X) = \mathcal{S}(G) - \mathcal{S}(X),$$

where G is a Gaussian vector in $\mathbb{R}^n$ with the same covariance matrix as X. This is an affine invariant quantity. In particular, if X is isotropic, its covariance matrix is the identity, thus $\mathcal{D}(X) = \frac{n}{2}\log(2\pi e) - \mathcal{S}(X)$ and we conclude that

$$\log\left(\sqrt{\frac{2\pi}{e}} L_X\right) \leq \frac{1}{n}\mathcal{D}(X) \leq \log\left(\sqrt{2\pi e} L_X\right).$$

The gain is that this inequality is affine invariant. We have thus established that the slicing problem is equivalent to: *for every $n \geq 1$ and every log-concave random vector X in $\mathbb{R}^n$, its relative entropy per coordinate $\frac{1}{n}\mathcal{D}(X)$ is bounded above by a universal constant.*

9 Coverings with translates

The idea of a covering is one of the simplest and most fundamental in mathematics. We shall be concerned with covering one convex set (particularly, the whole space) with translates of another one. Formally, for two convex sets K, L in $\mathbb{R}^n$ with nonempty interior, we define the **covering number**

$$\begin{aligned} N(K,L) &= \text{ smallest number of translates of } L \text{ needed to cover } K \\ &= \sup \left\{ m \geq 0,\ \exists\ x_1, \ldots, x_m \in \mathbb{R}^n, K \subset \bigcup_{i=1}^{m} (L + x_i) \right\}. \end{aligned}$$

Clearly, this is an affine invariant quantity. Of course, in many cases $N(\mathbb{R}^n, L) = \infty$.

To quantify efficiency of a covering of the whole space, we introduce the **covering density** $\vartheta(K)$ of a convex body K in $\mathbb{R}^n$ defined as

$$\vartheta(K) = \inf_{\mathcal{F}} \left\{ \limsup_{\substack{Q - \text{cube} \\ |Q| \to \infty}} \frac{1}{|Q|} \sum_{i: K_i \cap Q \neq \varnothing} |K_i| \right\},$$

where the infimum is taken over all coverings $\mathcal{F} = \{K_i,\ K_i = K + v_i, v_i \in \mathbb{R}^n\}$ of $\mathbb{R}^n$ by translates $K_i = K + v_i$ of K, that is $\mathbb{R}^n = \bigcup_{i=1}^{\infty} K_i$. Here are throughout this section, by a cube we mean a set of the form $[-s,s]^n + a$, $s > 0$, $a \in \mathbb{R}^n$.

It takes some effort to establish that $\vartheta(K)$ is also an affine invariant quantity (see, for instance Chapter 1.2 in [6]).

Given a convex body K in $\mathbb{R}^n$, we are interested in the quantity

$$b(K) = N(K, \text{int}K),$$

the minimum number of translates of the interior of K required to cover K. It is an exercise to show that it coincides with the minimum number of translates of smaller dilates λK, $0 < \lambda < 1$ of K required to cover K. Surprisingly, it also coincides with the minimum number of light sources required to illuminate the boundary of K (for these classical facts, see Paragraph 34 in [3]).

For instance, when K is a square on the plane, $b(K) = 4$, because any translate of its interior contains at most one vertex. It was shown that on the plane, $b(K) = 3$, unless K is an affine image of the cube. Similarly, when K is a cube in $\mathbb{R}^n$, we have $b(K) = 2^n$. It is conjectured that this is the worst case.

9.1 Conjecture (Gohberg, Marcus, Hadwiger, Levy)**.** For a convex body K in $\mathbb{R}^n$, $b(K) \leq 2^n$, with equality if and only if K is an affine image of the cube.

Our goal is to show the classical general result of Rogers which says that

$$b(K) \leq \binom{2n}{n}(n \log n + n \log\log n + 3n + 1)$$

for a convex body K in $\mathbb{R}^n$, $n \geq 3$. If K is symmetric, this can be significantly improved to

$$b(K) \leq 2^n(n \log n + n \log\log n + 3n + 1).$$

The proof is based on three results:

1) a general upper bound for covering numbers due to Rogers and Zong,

$$N(K, L) \leq \frac{|K - L|}{|K|}\vartheta(L),$$

2) a general upper bound for covering density due to Rogers,

$$\vartheta(K) \leq n \log n + n \log\log n + 3n + 1, \qquad n \geq 3,$$

3) an upper bound for the volume of the difference body due to Rogers and Shephard,

$$|K - K| \leq \binom{2n}{n}|K|.$$

9.1 A general upper bound for covering numbers

9.2 Theorem (Rogers and Zong)**.** *For convex subsets K, L of $\mathbb{R}^n$ with nonempty interior, we have*

$$N(K, L) \leq \frac{|K - L|}{|K|}\vartheta(L). \tag{9.1}$$

Proof. Fix $\varepsilon > 0$. By the definition of $\vartheta(L)$, there is a discrete subset G of $\mathbb{R}^n$ such that the translates of L by the elements of G cover $\mathbb{R}^n$, $\mathbb{R}^n = \bigcup_{g \in G}(L + g)$ and for every large enough cube Q,

$$\frac{1}{|Q|}\sum_{g \in G : (L+g) \cap Q \neq \varnothing} |L + g| \leq \vartheta(K) + \varepsilon$$

and since $L + g$ intersect Q if and only if $g \in Q - L$, we get

$$\#(G \cap (Q - L)) \leq \frac{|Q|}{|L|}(\vartheta(L) + \varepsilon). \tag{9.2}$$

Fix $t \in \mathbb{R}^n$. Since $L + g$ intersect $K + t$ if and only if $g \in K - L + t$ and since $\{L + g\}_{g \in G}$ covers $\mathbb{R}^n$, we conclude that $\{L + g\}_{g \in K - L + t}$ covers $K + t$. Let $N(t)$ be the size (cardinality) of this cover,

$$N(t) = \#(G \cap K - L + t) = \sum_{g \in G} \mathbf{1}_{K - L + t}(g).$$

Consider a large enough cube such that $K \subset \varepsilon Q$. Averaging $N(t)$ over Q gives

$$\frac{1}{|Q|}\int_Q N(t)\mathrm{d}t = \frac{1}{|Q|}\sum_{g \in G}\int_Q \mathbf{1}_{K-L+t}(g)\mathrm{d}t = \frac{1}{|Q|}\sum_{g \in G}|Q \cap (L - K + g)|.$$

Note that $|Q \cap (L - K + g)|$ equals 0 unless $g \in G \cap (K - L + Q)$ in which case it can be bounded by $|L - K + g| = |K - L|$. Thus,

$$\frac{1}{|Q|}\int_Q N(t)\mathrm{d}t \le \frac{|K-L|}{|Q|}\#(G \cap (K - L + Q)) \le \frac{|K-L|}{|Q|}\#(G \cap ((1+\varepsilon)Q - L)).$$

By (9.2) applied to $(1+\varepsilon)Q$, we get

$$\frac{1}{|Q|}\int_Q N(t)\mathrm{d}t \le (1+\varepsilon)^n \frac{|K-L|}{|L|}(\vartheta(L) + \varepsilon).$$

Therefore, there exists $t_0 \in Q$ such that $N(t_0)$ is bounded by the above quantity. This gives a covering of K of size $N(t_0)$ by sets $L + g - t_0$, hence

$$N(K, L) \le (1+\varepsilon)^n \frac{|K-L|}{|L|}(\vartheta(L) + \varepsilon).$$

Sending $\varepsilon \to 0$ finishes the proof. □

9.2 An asymptotic upper bound for covering densities

9.3 Theorem (Rogers)**.** *For a convex subset K of $\mathbb{R}^n$ with nonempty interior, $n \ge 3$, we have*

$$\vartheta(K) \le n \log n + n \log\log n + 3n + 1. \tag{9.3}$$

We shall need a simple observation about centred convex bodies.

9.4 Lemma. *If K is a centred convex body in $\mathbb{R}^n$, that is $\int_K x\mathrm{d}x = 0$, then $-\frac{1}{n}K \subset K$.*

Proof. Fix a direction $\theta \in S^{n-1}$ and consider the volume distribution function along this direction $f(t) = |K \cap (\theta^\perp + t\theta)|$. Brunn's principle (see Remark 2.4) says that f is $\frac{1}{n-1}$ concave on its support. Let $[-a, b]$, $a, b > 0$ be the support of f. Since K is centred, $\int_{-a}^b tf(t)\mathrm{d}t = 0$. Our goal is to show that $b \ge \frac{1}{n}a$ and $a \ge \frac{1}{n}b$. Let $w = f(0)^{\frac{1}{n-1}}$. Consider a linear function $g(x) = w\frac{x+a}{a}$. It agrees with $f^{\frac{1}{n-1}}$ at $x = -a$ and $x = 0$. By concavity, $g \le f^{\frac{1}{n-1}}$ on $[-a, 0]$ and $g \ge f^{\frac{1}{n-1}}$ on $[0, b]$. Thus

$$\int_0^b tg(t)^{n-1}\mathrm{d}t \ge \int_0^b tf(t)\mathrm{d}t = \int_{-a}^0 (-t)f(t)\mathrm{d}t \ge \int_{-a}^0 (-t)g(t)^{n-1}\mathrm{d}t.$$

Computing the left and right hand sides yields

$$w^{n-1}a^2\left(\frac{1}{n+1}\left[\left(1+\frac{b}{a}\right)^{n+1} - 1\right] - \frac{1}{n}\left[\left(1+\frac{b}{a}\right)^n - 1\right]\right) \ge w^{n-1}a^2\left(\frac{1}{n} - \frac{1}{n+1}\right)$$

which gives $b \ge \frac{1}{n}a$. A similar argument shows that $a \ge \frac{1}{n}b$. □

9.5 Remark. The proof shows that this result is tight for a cone (for one direction, $f^{\frac{1}{n-1}}$ is linear).

Proof of Theorem 9.3. Since $\vartheta(K)$ is affine invariant, we can assume that the volume of K is 1, $|K| = 1$ and the barycentre is at the origin, $\int_K x\mathrm{d}x = 0$.

Let $R > 0$ and consider the rescaled integer lattice $\Lambda = R\mathbb{Z}^n$. Taking R sufficiently large, we can ensure that every two translates of K by distinct $g_1, g_2 \in \Lambda$ are disjoint. Let $Q = [0, R]^n$. Let $X_1, \ldots, X_N$ be i.i.d. random vectors uniformly distributed on Q. Consider a (random) family of translates of K,

$$\mathcal{F} = \{K + X_i + g\}_{i \le N, g \in \Lambda}.$$

By a simple probabilistic argument (a first moment calculation), we shall show existence of the X_i such that $\mathcal{F}$ covers a *large portion* of the whole space. Let $E = \mathbb{R}^n \setminus \bigcup_{L \in \mathcal{F}} L$ be the set of uncovered points. Clearly,

$$\mathbf{1}_E(x) = \prod_{i=1}^N \big(1 - \mathbf{1}_{K+X_i+\Lambda}(x)\big).$$

Since for a fixed $i \le N$, the sets $K + X_i + g$, $g \in \Lambda$, are disjoint, we have

$$\mathbf{1}_{K+X_i+\Lambda}(x) = \sum_{g \in \Lambda} \mathbf{1}_{K+X_i+g}(x),$$

so

$$\mathbf{1}_E(x) = \prod_{i=1}^N \left(1 - \sum_{g \in \Lambda} \mathbf{1}_{K+X_i+g}(x)\right).$$

Consider

$$\rho = \frac{|E \cap Q|}{|Q|}$$

(since $\mathbf{1}_E$ is a periodic function in each coordinate with period R, ρ can be viewed as the density of the uncovered points). Let us compute its expectation. By independence, we have

$$\mathbb{E}\rho = \mathbb{E}\frac{1}{|Q|}\int_Q \mathbf{1}_E(x)\mathrm{d}x = \frac{1}{|Q|}\int_Q \prod_{i=1}^N \left(1 - \sum_{g \in \Lambda} \mathbb{E}\mathbf{1}_{K+X_i+g}(x)\right)\mathrm{d}x.$$

By the definition of X_i and a simple change of variables,

$$\mathbb{E}\mathbf{1}_{K+X_i+g}(x) = \frac{1}{|Q|}\int_Q \mathbf{1}_{K+u+g}(x)\mathrm{d}u = \frac{1}{|Q|}\int_{Q+g-x} \mathbf{1}_K(-v)\mathrm{d}v,$$

so using that the translates of Q by Λ are *almost* disjoint and cover $\mathbb{R}^n$,

$$\sum_{g \in \Lambda} \mathbb{E}\mathbf{1}_{K+X_i+g}(x) = \frac{1}{|Q|}\sum_{g \in \Lambda}\int_{Q+g-x} \mathbf{1}_K(-v)\mathrm{d}v = \frac{1}{|Q|}\int_{\mathbb{R}^n} \mathbf{1}_K(-v)\mathrm{d}v = \frac{|K|}{|Q|} = R^{-n}.$$

Thus,

$$\mathbb{E}\rho = \frac{1}{|Q|}\int_Q \big(1 - R^{-n}\big)^N = \big(1 - R^{-n}\big)^N.$$

Therefore, there exist points $X_1, \ldots, X_N \in Q$ for which

$$\rho \le \big(1 - R^{-n}\big)^N.$$

Fix such points and let $\mathcal{F} = \{K + X_i + g\}_{i \le N, g \in \Lambda}$. Summarising, this family covers the whole space but the set E and we have an upper bound for $\rho = \frac{|E \cap Q|}{|Q|}$.

The second part of the proof is to efficiently adjust $\mathcal{F}$ to a covering of the whole space. This is achieved by considering an appropriate maximal object. Fix $0 < \eta < \frac{1}{n}$. Consider points $y_1, \dots, y_M$ such that the family

$$\mathcal{G} = \{-\eta K + y_i + g\}_{i \le M, g \in \Lambda}$$

satisfies two conditions

(i) sets in $\mathcal{G}$ are disjoint,

(ii) sets in $\mathcal{G}$ do not intersect any set from $\mathcal{F}$.

By (i) and periodicity,

$$|Q \cap \bigcup_{L \in \mathcal{G}} L| = M\eta^n.$$

By (ii), every set from $\mathcal{G}$ is in E. Consequently,

$$|Q \cap \bigcup_{L \in \mathcal{G}} L| \le |Q \cap E|.$$

Let M be maximal such that points y_i exist and fix such points and $\mathcal{G}$ (if no y_i exists, $M = 0$). We thus have,

$$M = \eta^{-n}|Q \cap \bigcup_{L \in \mathcal{G}} L| \le \eta^{-n}|Q \cap E| = \eta^{-n}|Q|\rho \le \eta^{-n}R^n \left(1 - R^{-n}\right)^N.$$

Fix $z \in \mathbb{R}^n$. The sets $-\eta K + z + g$, $g \in \Lambda$ are disjoint, so since M is maximal, there is $g_1 \in \Lambda$ such that either

1) $-\eta K + z + g_1$ intersects a set from $\mathcal{G}$ (to violate (i)), or

2) $-\eta K + z + g_1$ intersects a sets from $\mathcal{F}$ (to violate (ii)).

In case 1), there are $x_1, x_2 \in K$, $i \le M$, $g_2 \in \Lambda$ such that $-\eta x_1 + z + g_1 = -\eta x_2 + y_i + g_2$, so, using Lemma 9.4, $-\eta K \subset K$,

$$z \in \eta(K - K) + y_i + (g_2 - g_1) \subset (1 + \eta)K + y_i + (g_2 - g_1).$$

In case 2), there are $x_1, x_2 \in K$, $i \le N$, $g_2 \in \Lambda$ such that $-\eta x_1 + z + g_1 = x_2 + X_i + g_2$, so

$$z \in (1 + \eta)K + X_i + (g_2 - g_1).$$

Since z is arbitrary, this shows that $\mathbb{R}^n$ is covered by

$$\{(1 + \eta)K + y_i + g\}_{i \le M, g \in \Lambda} \cup \{(1 + \eta)K + X_i + g\}_{i \le N, g \in \Lambda},$$

or, equivalently, by

$$\{K + (1 + \eta)^{-1}(y_i + g)\}_{i \le M, g \in \Lambda} \cup \{K + (1 + \eta)^{-1}(X_i + g)\}_{i \le N, g \in \Lambda}.$$

By periodicity, the density of this covering equals $(1+\eta)^n(M+N)R^{-n}$. Thus,

$$\vartheta(K) \le (1+\eta)^n(M+N)R^{-n}.$$

Using our bound for M, we obtain

$$\vartheta(K) \le (1+\eta)^n\eta^{-n}\left(1-R^{-n}\right)^N + (1+\eta)^nR^{-n}N.$$

The last part of the proof is to optimise over the parameters R, N and η. Regard η as fixed and take $R = \left(\frac{N}{n\log\frac{1}{\eta}}\right)^n$ with N sufficiently large such that R is large enough (as required at the beginning of the proof). Then $R^{-n} = \frac{n\log\frac{1}{\eta}}{N}$ and using $(1-R^{-n})^N \le e^{-NR^{-n}} = \eta^n$, we get

$$\vartheta(K) \le (1+\eta)^n + (1+\eta)^n n\log\frac{1}{\eta}.$$

For simplicity take $\eta = \frac{1}{n\log n}$ to obtain

$$\vartheta(K) \le \left(1+\frac{1}{n\log n}\right)^n (1+n\log n + n\log\log n).$$

Using $\left(1+\frac{1}{n\log n}\right)^n \le e^{\frac{1}{\log n}} \le 1+\frac{2}{\log n}$ for $n \ge 3$ and checking that $\frac{2}{\log n}(1+n\log n + n\log\log n) < 3n$ for $n \ge 3$ finishes the proof. □

9.3 An upper bound for the volume of the difference body

(This subsection was a guest lecture by B-H. Vritsiou.)

9.6 Theorem (Rogers-Shephard). *For a convex subset K of $\mathbb{R}^n$, we have*

$$|K-K| \le \binom{2n}{n}|K|. \tag{9.4}$$

Proof. Define $f(x) = |K\cap(K+x)|^{1/n}$ which is concave on its support $K-K$ (which follows from the Brunn-Minkowski inequality and a simple inclusion $K\cap(\lambda x+(1-\lambda)y) \supset \lambda(K\cap(L+x))+(1-\lambda)(K\cap(L+y))$ for arbitrary convex bodies K, L in $\mathbb{R}^n$, $\lambda\in[0,1]$ and $x,y\in\mathbb{R}^n$). For $x\in\mathbb{R}^n$ written in polar coordinates as $x = r\theta$, $r\ge 0$, $\theta\in S^{n-1}$, consider a function $g:\mathbb{R}^n\to\mathbb{R}$ defined as

$$g(r\theta) = f(0)\left(1-\frac{r}{\rho_{K-K}(\theta)}\right),$$

where $\rho_{K-K}(\theta) = \sup\{t>0, t\theta\in K-K\}$ is the radial function of $K-K$ in the direction θ. Note that along each ray $\{t\theta, t\ge 0\}$, f is concave, g is linear agreeing with f at $t=0$ and $t=\rho_{K-K}(\theta)$. Thus, $f\ge g$ on every segment $[0,\theta\rho_{K-K}(\theta)]$ and therefore $f\ge g$ on $K-K$. From this we obtain

$$\int_{K-K} f^n \ge \int_{K-K} g^n$$

and the right hand side can be computed using polar coordinates as follows

$$\begin{aligned}\int_{K-K} g^n &= f(0)^n \int_{S^{n-1}} \int_0^{\rho_{K-K}(\theta)} \left(1 - \frac{r}{\rho_{K-K}(\theta)}\right)^n r^{n-1}|S^{n-1}|\mathrm{d}r\mathrm{d}\sigma(\theta) \\ &= f(0)^n \int_{S^{n-1}} \int_0^1 (1-t)^n t^{n-1}|S^{n-1}|\rho_{K-K}(\theta)^n \mathrm{d}t\mathrm{d}\sigma(\theta) \\ &= f(0)^n \left(|S^{n-1}| \int_{S^{n-1}} |S^{n-1}|\rho_{K-K}(\theta)^n \mathrm{d}\sigma(\theta)\right)\left(\int_0^1 t^{n-1}(1-t)^n \mathrm{d}t\right).\end{aligned}$$

By the formula for the volume in polar coordinates, the first parenthesis equals $n|K-K|$. The second one is $B(n, n+1) = \frac{\Gamma(n)\Gamma(n+1)}{\Gamma(2n+1)} = \frac{(n-1)!n!}{(2n)!} = \frac{1}{n\binom{2n}{n}}$. Recall the definition of f to see that $f(0)^n = |K|$ and conclude

$$\int_{K-K} f^n \geq \int_{K-K} g^n = |K|\frac{|K-K|}{\binom{2n}{n}}.$$

On the other hand, by Fubini's theorem,

$$\begin{aligned}\int_{K-K} f^n = \int_{\mathbb{R}^n} |K \cap (K+x)|\mathrm{d}x &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \mathbf{1}_K(y)\mathbf{1}_{K+x}(y)\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}^n} \mathbf{1}_K(y)\left(\int_{\mathbb{R}^n} \mathbf{1}_{y-K}(x)\mathrm{d}x\right)\mathrm{d}y \\ &= \int_{\mathbb{R}^n} \mathbf{1}_K(y)|K|\mathrm{d}y = |K|^2.\end{aligned}$$

Putting the last two conclusions together finishes the proof. □

9.7 Remark. The same arguments give a more general inequality: for convex bodies K and L in $\mathbb{R}^n$, we have

$$|K+L| \leq \binom{2n}{n}\frac{|K|\cdot|L|}{|K\cap -L|} \tag{9.5}$$

(the point being that the function $f(x) = |K \cap (L+x)|^{1/n}$ is concave).

9.8 Remark. For convex bodies K and L in $\mathbb{R}^n$, we also have

$$|K-L| \leq \binom{2n}{n}|K+L|. \tag{9.6}$$

Assuming without loss of generality that 0 belongs to both K and L, we have an inclusion $K - L \subset K + L - (K + L)$, so (9.4) yields

$$|K-L| \leq |K+L-(K+L)| \leq \binom{2n}{n}|K+L|.$$

9.9 Remark. By the Brunn-Minkowski inequality $|K-K| \geq 2^n|K|$. Combining this with (9.4) and using $\binom{2n}{n} < 4^n$, we get that the volume of the symmetric difference $K - K$ is comparable to the volume of K on the exponential scale,

$$2 \leq \frac{|K-K|^{1/n}}{|K|^{1/n}} \leq 4.$$

9.10 Remark. Using equality cases for the Brunn-Minkowski and deriving a nontrivial characterisation of a simplex, Rogers and Shephard also showed that (9.4) becomes equality if and only if K is a simplex.

A Appendix: Haar measure

We begin with recalling an abstract theorem guaranteeing existence of Haar measure.

A.1 Theorem. *Let (M, d) be a compact metric space and let G be a group acting on M as isometries, that is $d(gx, gy) = d(x, y)$ for $x, y \in M$ and $g \in G$. There exists a regular finite Borel measure μ on M which is invariant under the action of G, that is $\mu(gA) = \mu(A)$ for all $g \in G$ and Borel subsets A of M. Moreover, μ is unique up to a constant, if the action of G on M is transitive (for every $x, y \in M$, there is $g \in G$ such that $x = gy$).*

Such a measure is called a **Haar measure**. It is often normalised to be a probability measure and we shall make no exception. Let us discuss three important examples: the sphere, orthogonal group and Grassmannian.

Sphere

Consider the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n,\ |x| = 1\}$ in $\mathbb{R}^n$. It is naturally equipped with the Euclidean metric: for $x, y \in S^{n-1}$, $d_E(x, y) = |x - y|$. There is also the **geodesic metric** $d_G(x, y)$ defined as the measure of the convex angle $x0y$ (on the plane spanned by x and y). We can check that $|x - y| = 2\sin\frac{d_G(x,y)}{2}$, thus $\frac{2}{\pi}d_G \leq d_E \leq d_G$. The orthogonal group $O(n)$ acts transitively on S^{n-1} as isometries. The unique probability Haar measure σ on S^{n-1} provided by Theorem A.1 is the normalised surface (Lebesgue) measure on S^{n-1} which also agrees with its cone measure, that is

$$\sigma(A) = \frac{|\mathrm{cone}(A)|}{|B_2^n|},$$

for a Borel subset A of S^{n-1}, where $\mathrm{cone}(A) = \{ta, t \in [0, 1], a \in A\}$. These two statements are justified by the invariance and uniqueness properties of the Haar measure.

The Haar measure on the sphere is related to the standard Gaussian measure by the following extremely useful factorisation result, which is very intuitive.

A.2 Theorem. *Let G be a standard Gaussian vector in $\mathbb{R}^n$. Take Θ to be a random vector uniformly distributed on the unit sphere S^{n-1} and R to be an independent non-negative random variable with density $\frac{|S^{n-1}|}{\sqrt{2\pi}^n}r^{n-1}e^{-r^2/2}$ on $[0, \infty)$. Then G has the same distribution as $R \cdot \Theta$, $G \overset{d}{=} R \cdot \Theta$.*

Proof. Integrating in spherical coordinates, for a measurable function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$\begin{aligned}\mathbb{E}f(G) = \int_{\mathbb{R}^n} f(x)e^{-|x|^2/2}\frac{\mathrm{d}x}{\sqrt{2\pi}^n} &= \int_{S^{n-1}}\int_0^\infty f(r\theta)e^{-r^2/2}r^{n-1}\frac{|S^{n-1}|}{\sqrt{2\pi}^n}\mathrm{d}\sigma(\theta)\mathrm{d}r \\ &= \int_0^\infty \left(\int_{S^{n-1}} f(r\theta)\mathrm{d}\sigma(\theta)\right)\frac{|S^{n-1}|}{\sqrt{2\pi}^n}r^{n-1}e^{-r^2/2}\mathrm{d}r \\ &= \mathbb{E}_R\mathbb{E}_\Theta f(R\Theta) = \mathbb{E}f(R\Theta).\end{aligned}$$

□

In particular, this allows to compute Gaussian Euclidean moments. For $p > -n$, we have

$$\mathbb{E}|G|^p = \mathbb{E}|R\Theta|^p = \mathbb{E}R^p = \int_0^\infty r^{n-1+p} e^{-r^2/2} \frac{|S^{n-1}|}{\sqrt{2\pi}^n} \mathrm{d}r.$$

Changing variables and rearranging yields,

$$\mathbb{E}|G|^p = 2^{\frac{n+p}{2}-1}\Gamma\left(\frac{n+p}{2}\right)\frac{|S^{n-1}|}{\sqrt{2\pi}^n}.$$

Setting $p = 0$ gives $\frac{|S^{n-1}|}{\sqrt{2\pi}^n} = 2^{-n/2+1}/\Gamma(n/2)$ and we get

$$\mathbb{E}|G|^p = \frac{2^{\frac{n}{2}}\Gamma\left(\frac{n+p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \tag{A.1}$$

In particular, for $p = 1$, thanks to Stirling's formula $\Gamma(x) \sim \sqrt{2\pi}x^{x-\frac{1}{2}}e^{-x}$,

$$\mathbb{E}|G| = \mathbb{E}R = \frac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sim \frac{\sqrt{2}\left(\frac{n+1}{2}\right)^{\frac{n}{2}}}{\left(\frac{n}{2}\right)^{\frac{n}{2}-\frac{1}{2}}} e^{-\frac{1}{2}} = e^{-1/2}\left(1+\frac{1}{n}\right)^{n/2}\sqrt{n}$$

and we obtain

$$\mathbb{E}|G| = \mathbb{E}R = (1+o(1))\sqrt{n}. \tag{A.2}$$

Consequently, writing $G \stackrel{d}{=} R\Theta$, for any norm $\|\cdot\|$ on $\mathbb{R}^n$, we have

$$\mathbb{E}\|G\| = (1+o(1))\sqrt{n}\mathbb{E}\|\Theta\|. \tag{A.3}$$

Orthogonal group

Consider the **orthogonal group** $O(n)$ (all orthogonal $n \times n$ real matrices). It can be equipped for instannce with the operator norm and then it acts on itself as isometries. The Haar measure ν_n on $O(n)$ thus satisfies $\nu_n(UAV) = \nu_n(A)$ for all Borel subsets A of $O(n)$ and $U, V \in O(n)$. Practically, Haar measure ν_n can be realised as follows: take a random vector uniformly distributed on the unit sphere, then take a random vector θ_2 uniformly distributed on the unit sphere conditioned on being perpendicular to θ_1, then take a random vector θ_3 uniformly distributed on the unit sphere conditioned on being perpendicular to θ_1 and θ_2, etc. Then the random matrix whose columns are $\theta_1, \ldots, \theta_n$ is distributed according to ν_n.

The Haar measures σ on S^{n-1} and ν_n on $O(n)$ are of course related: thanks to invariance and uniqueness, for a Borel set A in S^{n-1} and a unit vector $x \in S^{n-1}$, we have

$$\nu_n(\{U \in O(n),\ Ux \in A\}) = \sigma(A). \tag{A.4}$$

In other words, if U is a uniform random matrix on $O(n)$ and θ is a uniform random vector on S^{n-1}, then for any (fixed) vector $x \in \mathbb{R}^n$,

$$Ux \stackrel{d}{=} |x|\theta. \tag{A.5}$$

Grassmannian

Consider the **Grassmannian** $G_{n,k}$, that is the set of all k-dimensional subspaces in $\mathbb{R}^n$. It can be equipped for instance with the Hausdorff distance between the unit balls of two subspaces and then the orthogonal group acts on the Grassmannian as isometries. The Haar measure $\nu_{n,k}$ on $G_{n,k}$ satisfies $\nu_{n,k}(UA) = \nu_{n,k}(A)$, for $U \in O(n)$, and Borel sets $A \subset G_{n,k}$. A way to generate $\nu_{n,k}$ is of course to use the Haar measure on the orthogonal group: let U be a uniform random matrix on $O(n)$ and let F be a fixed subspace in $\mathbb{R}^n$ of dimension k; then UF is a uniform random subspace in $G_{n,k}$, that is for any Borel subset A of $G_{n,k}$,

$$\nu_n(\{U \in O(n),\ UF \in A\}) = \nu_{n,k}(A). \tag{A.6}$$

We conclude with the following useful decomposition identity: for an integrable function $f : S^{n-1} \to \mathbb{R}$, we have

$$\int_{S^{n-1}} f \mathrm{d}\sigma = \int_{G_{n,k}} \left(\int_{S_F} f \mathrm{d}\sigma_F \right) \mathrm{d}\nu_{n,k}(F), \tag{A.7}$$

where $S_F = S^{n-1} \cap F$ is the unit sphere in F and σ_F is its Haar measure. As always, both (A.6) and (A.7) can be checked using invariance and uniqueness (for the latter it helps check it first for indicators).

B Appendix: Spherical caps

A **spherical cap** on the unit sphere S^{n-1}, centred at $\theta \in S^{n-1}$ with radius $r > 0$, equivalently, distance $\varepsilon = 1 - \frac{r^2}{2}$ from the origin (height $1 - \varepsilon$) is the set

$$\begin{aligned} C(\theta, \varepsilon) &= \{x \in S^{n-1},\ \langle x, \theta \rangle \geq \varepsilon\} \\ &= \{x \in S^{n-1},\ |x - \theta| \leq r\} = B(\theta, r). \end{aligned}$$

It is useful to have a good bound for the measure of spherical caps. In the next two theorems, we provide simple upper and lower bounds.

B.1 Theorem. *For $\varepsilon \in [0, 1]$ and $\theta \in S^{n-1}$, we have*

$$\sigma(C(\theta, \varepsilon)) \leq e^{-\varepsilon^2 n/2}.$$

Proof. We shall use the cone measure representation for σ. Let A be the cone based on $C(\theta, \varepsilon)$ intersected with B_2^n. We distinguish two cases.
Case 1. If $0 \leq \varepsilon \leq \frac{1}{\sqrt{2}}$, then $A \subset \varepsilon\theta + \sqrt{1 - \varepsilon^2} B_2^n$, thus

$$\sigma(C(\theta, \varepsilon)) = \frac{|A|}{|B_2^n|} \leq \frac{|\sqrt{1 - \varepsilon^2} B_2^n|}{|B_2^n|} = (1 - \varepsilon^2)^{n/2} \leq e^{-\varepsilon^2 n/2}.$$

Case 2. If $\frac{1}{\sqrt{2}} \leq \varepsilon \leq 1$, then $A \subset \frac{1}{2\varepsilon}\theta + \frac{1}{2\varepsilon}B_2^n$, thus

$$\sigma(C(\theta,\varepsilon)) = \frac{|A|}{|B_2^n|} \leq \frac{|\frac{1}{2\varepsilon}B_2^n|}{|B_2^n|} = \left(\frac{1}{2\varepsilon}\right)^n \leq e^{-\varepsilon^2 n/2}.$$

The last estimate follows from the inequality $e^{x^2/2} < 2x$, $x \in [\frac{1}{\sqrt{2}}, 1]$, which, by convexity, reduces to verifying it at $x = \frac{1}{\sqrt{2}}$ and $x = 1$. □

B.2 Theorem. *For $r \in [0,2]$ and $\theta \in S^{n-1}$, we have*

$$\sigma(B(\theta,r)) \geq \left(\frac{r}{4}\right)^n.$$

Proof. Let X be an r-net in S^{n-1} of size at most $(1+2/r)^n$ (see below). Thus,

$$1 = \sigma(S^{n-1}) \leq \sigma\left(\bigcup_{\theta \in X} B(\theta,r)\right) \leq |X| \cdot \sigma(B(\theta,r)),$$

consequently,

$$\sigma(B(\theta,r)) \geq \left(\frac{r}{r+2}\right)^n \geq \left(\frac{r}{4}\right)^n.$$

□

For the convenience of our proofs, we stated the above upper and lower bounds using two different parametrisations of caps, but of course, we can easily translate one into another using $\varepsilon = 1 - \frac{r^2}{2}$.

We finish by explaining the existence of small nets, which is a very useful fact (beyond the application we just saw in Theorem B.2). Recall that a δ-net of a metric space (M,d) is a subset X of M such that that for every point y from M, there is a point x in X such that $d(x,y) < \delta$. In other words, M is covered with the balls with radius r centred at the points in X, $M \subset \bigcup_{x \in X} B(x,\delta)$.

B.3 Lemma. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. For every $\delta > 0$, there is a δ-net with respect to the distance measured by $\|\cdot\|$ of its unit sphere $\{x \in \mathbb{R}^n,\ \|x\| = 1\}$ of size at most $(1+2/\delta)^n$.*

Proof. Let $B = \{x \in \mathbb{R}^n,\ \|x\| < 1\}$ be the unit ball and let $S = \{x \in \mathbb{R}^n,\ \|x\| = 1\}$ be the unit sphere with respect to $\|\cdot\|$. Let X be a subset of S of maximal cardinality with the property that every two points of X are at least δ-apart in distance measured by $\|\cdot\|$, equivalently, the balls $\{x + \frac{\delta}{2}B\}_{x \in X}$ are disjoint. Note that by its maximality, X is also a δ-net of S (otherwise, we could add a point to X). By a volume argument, X cannot be too large,

$$|X| \cdot (\delta/2)^n \operatorname{vol}_n(B) = \operatorname{vol}_n\left(\bigcup_{x \in X}(x + \frac{\delta}{2}B)\right) \leq \operatorname{vol}_n\left(\left(1 + \frac{\delta}{2}\right)B\right)$$
$$= \left(1 + \frac{\delta}{2}\right)^n \operatorname{vol}_n(B),$$

hence $|X| \leq (1+2/\delta)^n$. □

C Appendix: Stirling's Formula for Γ

Recall that Stirling's formula for factorials of integers says that

$$n! = \sqrt{2\pi} n^{n+1/2} e^{-n} \left(1 + O\left(\frac{1}{n}\right)\right), \qquad n \to \infty. \tag{C.1}$$

This extends to the continuous case when we consider the Gamma function

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

We have $\Gamma(x+1) = x\Gamma(x)$ and thus, for integers, $\Gamma(n+1) = n!$. Stirling's formula reads

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x} \left(1 + O\left(\frac{1}{x}\right)\right), \qquad x \to \infty. \tag{C.2}$$

To recover (C.1), set $x = n$ and multiply both sides of (C.2) by n. In fact precise two sided bounds are known.

C.1 Theorem. *For $x > 0$, we have*

$$\sqrt{2\pi} x^{x-1/2} e^{-x} \leq \Gamma(x) \leq \sqrt{2\pi} x^{x-1/2} e^{-x} e^{\frac{1}{12x}}. \tag{C.3}$$

A complete proof with a discussion and other references can be found in [5].

References

[1] Artstein-Avidan, S., Giannopoulos, A., Milman, V., Asymptotic geometric analysis. Part I. Providence, RI, 2015.

[2] Ball, K., An elementary introduction to modern convex geometry. Cambridge, 1997.

[3] Boltyanski, V., Martini, H., Soltan, P. S., Excursions into combinatorial geometry. Universitext. *Springer-Verlag, Berlin*, 1997.

[4] Brazitikos, S., Giannopoulos, A., Valettas, P., Vritsiou, B., Geometry of isotropic convex bodies. Providence, RI, 2014.

[5] Jameson, G., A simple proof of Stirling's formula for the gamma function. *Math. Gaz.* 99 (2015), no. 544, 6874.

[6] Rogers, C. A., Packing and covering. Cambridge Tracts in Mathematics and Mathematical Physics, No. 54 *Cambridge University Press, New York*, 1964.