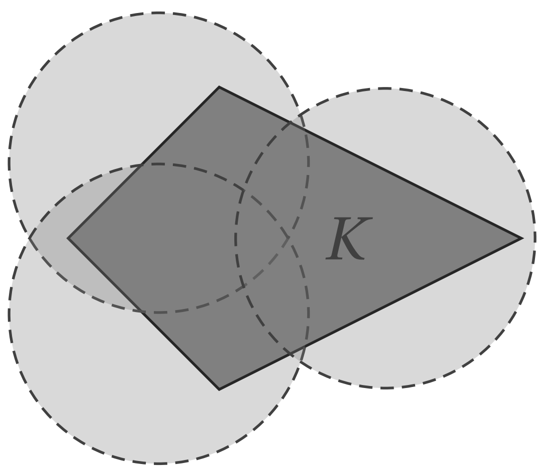
Asymptotic Convex Geometry

Winter 2021

Thomas Rothvoss



UNIVERSITY *of* WASHINGTON

Last changes: February 11, 2024

Contents

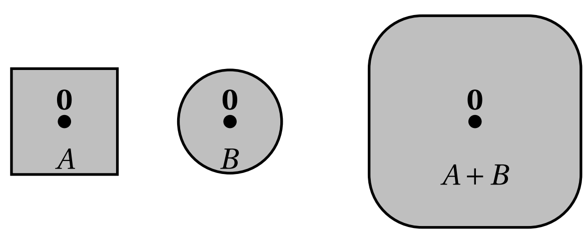
Chapter 1

Basics of Convex Geometry

The goal of this monograph is a introduction to high-dimensional convex geometry. As a basis we follow a selection of chapters of the excellent textbook *Asymptotic Geometric Analysis, Part 1* by Artstein-Avidan, Giannopolous and Milman [AAGM15]. However, in presentation we try a more geometric rather than functional-analytic approach. We will avoid working with infinite-dimensional vector spaces where ever possible. We will also make use of simplified proofs if available. I am in particular grateful to Victor Reis for proof reading this manuscript.

1.1 Basic Definitions

A set $A \subseteq \mathbb{R}^n$ is *convex* if $(1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y} \in A$ for all $\boldsymbol{x}, \boldsymbol{y} \in A$ and $0 \leq \lambda \leq 1$. Let $B_2^n := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 \leq 1\}$ be the *Euclidean ball* of radius 1 around the origin. Moreover, we define $S^{n-1} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 = 1\}$ as the $(n-1)$*-dimensional sphere*. More generally, we define the balls $B_p^n := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_p \leq 1\}$ where $\|\boldsymbol{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $1 \leq p < \infty$ and $\|\boldsymbol{x}\|_\infty := \max_{i=1,\ldots,n} |x_i|$. The *Minkowski sum* of two sets $A, B \subseteq \mathbb{R}^n$ is defined by $A + B := \{\boldsymbol{a} + \boldsymbol{b} \mid \boldsymbol{a} \in A, \boldsymbol{b} \in B\}$.

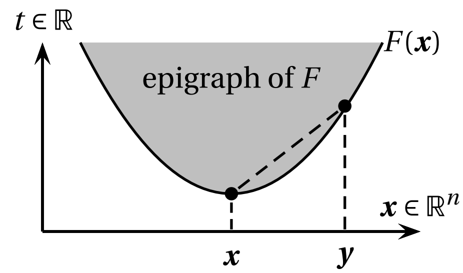


We denote $\text{int}(K) := \{\boldsymbol{x} \in K \mid \exists \varepsilon > 0 : \boldsymbol{x} + \varepsilon B_2^n \subseteq K\}$ as the *interior* of K. A *convex body* is a convex set that is compact (= closed and bounded) and has a non-empty interior. We say that K is *(centrally) symmetric* if $K = -K$.

A function $F : \mathbb{R}^n \to \mathbb{R}$ is *convex* if

$$F((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{y}) \leq (1-\lambda)F(\boldsymbol{x}) + \lambda F(\boldsymbol{y}) \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \ \forall 0 \leq \lambda \leq 1.$$

In other words for every $\boldsymbol{x}$ and $\boldsymbol{y}$, the line segment between $(\boldsymbol{x}, F(\boldsymbol{x}))$ and $(\boldsymbol{y}, F(\boldsymbol{y}))$ lies above the graph. This is equivalent to asking that the *epigraph* $\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid t \geq F(\boldsymbol{x})\}$ is convex.



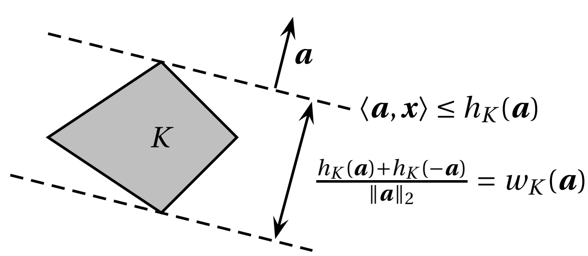
The *support function* for a set K is defined by

$$h_K(\boldsymbol{a}) := \sup_{\boldsymbol{x} \in K} \langle \boldsymbol{a}, \boldsymbol{x} \rangle$$

where $\boldsymbol{a} \in \mathbb{R}^n$. In other words, $h_K(\boldsymbol{a})$ is the minimal value so that $\langle \boldsymbol{a}, \boldsymbol{x} \rangle \leq h_K(\boldsymbol{a})$ is a valid inequality for K. If $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$, then

$$w_K(\boldsymbol{a}) := \frac{h_K(\boldsymbol{a}) + h_K(-\boldsymbol{a})}{\|\boldsymbol{a}\|_2} = \sup\left\{\frac{|\langle \boldsymbol{a}, \boldsymbol{x} \rangle - \langle \boldsymbol{a}, \boldsymbol{y} \rangle|}{\|\boldsymbol{a}\|_2} : \boldsymbol{x}, \boldsymbol{y} \in K\right\}$$

is the *width* of K in direction $\boldsymbol{a}$. Geometrically speaking, $w_K(\boldsymbol{a})$ is the minimal width of a strip with normal vector $\boldsymbol{a}$ that contains K. Note that for a convex body, the supremum is always attained and in the definitions of $h_K(\boldsymbol{a})$ and $w_K(\boldsymbol{a})$, we could replace the sup with a max.



Let $\mathrm{Vol}_n(K)$ be the *n-dimensional volume* of a body K. We will study also other measure for the "largeness" of a set. In particular for a non-empty convex set K we define the *mean width* as $w(K) := \mathbb{E}_{\boldsymbol{x} \sim S^{n-1}}[w_K(\boldsymbol{x})]$. Yet, a different measure is the *(standard) Gaussian measure* γ_n which is defined as

$$\gamma_n(K) := \int_K \frac{1}{(2\pi)^{n/2}} \cdot e^{-\|\boldsymbol{x}\|_2^2/2} \, d\boldsymbol{x}$$

By a slight abuse of notation, we will write $\gamma_n(\boldsymbol{x}) := \frac{1}{(2\pi)^{n/2}} \cdot e^{-\|\boldsymbol{x}\|_2^2/2}$ as the *density* of this distribution. The distribution has many exceptional properties: it is rotationally symmetric and drawing standard Gaussian $\boldsymbol{x} \sim \gamma_n$ is equivalent to drawing the coordinate entries $x_1, \dots, x_n \sim \gamma_1$ independently. Without making the statement formal at this point, the Gaussian measure $\gamma_n(K)$ is approximately the volume ratio $\mathrm{Vol}_n(K \cap \sqrt{n} B_2^n)/\mathrm{Vol}_n(B_2^n)$. A random vector $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ is called a *Gaussian random vector* if $\boldsymbol{x} \sim \gamma_n$ and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a matrix and $\boldsymbol{b} \in \mathbb{R}^n$. The Gaussian is centered if $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{b} = \boldsymbol{0}$. Gaussians are universal in the sense that adding independent random vectors together gives a distribution that converges to the Gaussian with identical expectation and covariance matrix.

Lemma 1.1. *The quantity $a_n := \mathbb{E}_{\boldsymbol{x} \sim \gamma_n}[\|\boldsymbol{x}\|_2]$ satisfies $\sqrt{n} \cdot \sqrt{\frac{n}{n+1}} \leq a_n \leq \sqrt{n}$.*

The upper bound follows from Jensen's inequality and $\mathbb{E}_{\boldsymbol{x} \sim \gamma_n}[\|\boldsymbol{x}\|_2^2] = n$ — we skip the lower bound calculations here. We will also later use the *Gaussian mean width*

$$g(K) := \mathop{\mathbb{E}}_{\boldsymbol{a} \sim \gamma_n} \left[\sup_{\boldsymbol{x} \in K} \langle \boldsymbol{a}, \boldsymbol{x} \rangle \right]$$

Note that in contrast to $w(K)$ this is a "one-sided" notion of width, but $g(K) = \frac{a_n}{2} w(K) \approx \frac{\sqrt{n}}{2} \cdot w(K)$ where the multiplicative error goes to 0 as $n \to \infty$.

1.2 Norms, Polarity and dual norms

If V is an $\mathbb{R}$-vector space (most of the time we will simply consider $V = \mathbb{R}^n$), then a map $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is called a *norm* if it satisfies

(i) subadditivity: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$
(ii) homogenity: $\|\lambda \boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|$ for $\lambda \in \mathbb{R}$ and $\boldsymbol{x} \in V$
(iii) point-separation: $\|\boldsymbol{x}\| = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$

If we have a symmetric convex body K, then *Minkowski norm* is defined as

$$\|\boldsymbol{x}\|_K := \min\{\lambda \geq 0 : \boldsymbol{x} \in \lambda K\}$$

Indeed, one can check that this is a norm as the convexity of K implies the subadditivity of $\|\cdot\|_K$ and the symmetry implies the homogenity. In fact, for any norm $\|\cdot\|$ we can set $K := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \leq 1\}$ as the unit ball of that norm, then $\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_K$ for any $\boldsymbol{x}$. For example one has $\|\boldsymbol{x}\|_p = \|\boldsymbol{x}\|_{B_p^n}$.

For $K \subseteq \mathbb{R}^n$, let $\mathrm{span}(K)$ be the unique minimal subspace with $K \subseteq \mathrm{span}(K)$. We now come to a very crucial concept in convex geometry:

Definition 1.2. For a convex set $K \subseteq \mathbb{R}^n$ with $\mathbf{0} \in K$ we define the *polar* as

$$K^\circ := \left\{ \boldsymbol{y} \in \operatorname{span}(K) \mid \sup\{\langle \boldsymbol{x}, \boldsymbol{y} \rangle : \boldsymbol{x} \in K\} \leq 1 \right\}$$

If K is a convex body with $\mathbf{0} \in K$ then the definition simplifies to

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

Recall that $K \subseteq \mathbb{R}^n$ is a *polytope* if $K = \operatorname{conv}(S)$ for a finite set of points S. Equivalently K is a polytope if K is bounded and has a finite number of faces, i.e. $K = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle \leq b_i \ \ \forall i = 1, \ldots, N\}$. Any convex body can be arbitrarily well approximated by a polytope. To gain intuition we discuss polarity for polytopes.

Lemma 1.3. *Let K be a polytope with $\mathbf{0} \in \operatorname{int}(K)$.*

(a) If $K = \operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ then $K^\circ = \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_1, \boldsymbol{y} \rangle \leq 1, \ldots, \langle \boldsymbol{a}_N, \boldsymbol{y} \rangle \leq 1\}$.
(b) If $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_1, \boldsymbol{x} \rangle \leq 1, \ldots, \langle \boldsymbol{a}_N, \boldsymbol{x} \rangle \leq 1\}$, then $K^\circ = \operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$.

Moreover, in both cases K° is again a polytope with $\mathbf{0} \in \operatorname{int}(K^\circ)$.

Proof. Note that the boundedness of K implies that $\mathbf{0} \in \operatorname{int}(K^\circ)$.

For (a) we simply use the definition to get

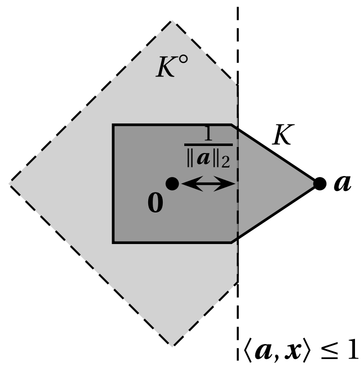
$$\begin{aligned} K^\circ &= \left\{ \boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{x} \rangle \leq 1 \quad \forall \boldsymbol{x} \in K \right\} \\ &= \left\{ \boldsymbol{y} \in \mathbb{R}^n : \sum_{i=1}^N \lambda_i \langle \boldsymbol{y}, \boldsymbol{a}_i \rangle \leq 1 \quad \forall \boldsymbol{\lambda} \in \mathbb{R}^N_{\geq 0} : \sum_{i=1}^N \lambda_i = 1 \right\} \\ &= \left\{ \boldsymbol{y} \in \mathbb{R}^n : \langle \boldsymbol{y}, \boldsymbol{a}_i \rangle \leq 1 \ \forall i \in [N] \right\} \end{aligned}$$

Here we have used that any point $\boldsymbol{y}$ that satisfies the N linear constraints will also satisfy any convex combination of them.

Now consider (b). First we prove $C := \operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} \subseteq K^\circ$. Let $\boldsymbol{\lambda} \in \mathbb{R}^N_{\geq 0}$ with $\sum_{i=1}^N \lambda_i = 1$ be a convex combination. We have $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle \leq 1 \ \forall \boldsymbol{x} \in K$ by assumption and so $\langle \sum_{i=1}^N \lambda_i \boldsymbol{a}_i, \boldsymbol{x} \rangle \leq 1$ for all $\boldsymbol{x} \in K$. This implies $\sum_{i=1}^N \lambda_i \boldsymbol{a}_i \in K^\circ$. If $\mathbf{0} \notin C$ then by the Hyperplane Separation Lemma there is a direction $\boldsymbol{x}$ with $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle \leq 0$ meaning that K is unbounded in direction $\boldsymbol{x}$. Hence $\mathbf{0} \in C$. Now consider a point $\boldsymbol{y} \notin C$. Again by the Hyperplane Separation Lemma, there is a normal vector $\boldsymbol{x} \in \mathbb{R}^n$ with $\langle \boldsymbol{a}, \boldsymbol{x} \rangle < \beta < \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{a} \in C$. As $\mathbf{0} \in C$ we know that $\beta > 0$. So after scaling $\boldsymbol{x}$ we may assume that $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle < 1 < \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $i \in [N]$. Then $\boldsymbol{x} \in K$ and this is a certificate that $\boldsymbol{y} \notin K^\circ$.

Part (b) also proves that the polar of any bounded polytope with $\mathbf{0} \in \operatorname{int}(K)$ is bounded, which settles the "moreover" part. $\square$

More intuitively, Lemma 1.3 shows that when moving from K to the polar K°, points turn to inequalities and inequalities turn to points. We would like to mention that the polar is not invariant under translation.



One has the following useful fact:

Theorem 1.4 (Polarity Theorem)**.** *For a convex body $K \subseteq \mathbb{R}^n$ with $\mathbf{0} \in int(K)$ one has (a) $(K^\circ)^\circ = K$ and (b) if K is also symmetric then $\|\boldsymbol{x}\|_K = h_{K^\circ}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Proof. We verify this at least for polytopes. Claim (a) follows from Lemma 1.3. For (b) we write the polytope in the form $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_1, \boldsymbol{x}\rangle \leq 1, \ldots, \langle \boldsymbol{a}_N, \boldsymbol{x}\rangle \leq 1\}$. Then we observe that

$$\|\boldsymbol{x}\|_K = \max\{\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle : i = 1, \ldots, N\}$$

while

$$h_{K^\circ}(\boldsymbol{x}) = \max\left\{\langle \boldsymbol{y}, \boldsymbol{x}\rangle \mid \boldsymbol{y} \in \mathrm{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}\right\} = \max\{\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle : i = 1, \ldots, N\}$$

$\square$

For a norm $\|\cdot\|$ in $\mathbb{R}^n$, we define the *dual norm* as

$$\|\boldsymbol{x}\|_* := \sup\left\{\langle \boldsymbol{y}, \boldsymbol{x}\rangle : \boldsymbol{y} \in \mathbb{R}^n \text{ with } \|\boldsymbol{y}\| \leq 1\right\}$$

From the definition we observe the following immediately:

Lemma 1.5. *Let K be a symmetric convex body. Then $\|\cdot\|_{K^\circ}$ is the dual norm of $\|\cdot\|_K$.*

Proof. It suffices to check that for $\|\cdot\| := \|\cdot\|_K$ one has

$$\begin{aligned} \|\boldsymbol{x}\|_* \leq 1 \quad &\Leftrightarrow \quad \langle \boldsymbol{y}, \boldsymbol{x}\rangle \leq 1 \; \forall \boldsymbol{y} \in \mathbb{R}^n \text{ with } \|\boldsymbol{y}\|_K \leq 1 \\ &\Leftrightarrow \quad \langle \boldsymbol{y}, \boldsymbol{x}\rangle \leq 1 \;\; \forall \boldsymbol{y} \in K \\ &\Leftrightarrow \quad \boldsymbol{x} \in K^\circ \end{aligned}$$

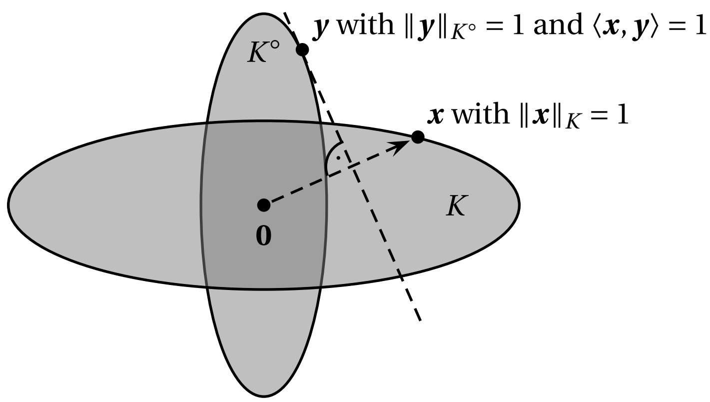
$\square$

Another immediate consequence of the definition of the dual norm:

Lemma 1.6 (Generalized Cauchy-Schwarz)**.** *For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and any symmetric convex set K one has $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|_K \cdot \|\boldsymbol{y}\|_{K^\circ}$.*

Proof. After scaling we may assume $\|\boldsymbol{x}\|_K = 1$. Then $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{y}\|_{K^\circ}$ by definition of dual norm. □

Yet another useful observation is the following: For any $\boldsymbol{x} \in \mathbb{R}^n$ and any norm $\|\cdot\|_K$ there must be an element $\boldsymbol{y}$ with $\|\boldsymbol{y}\|_{K^\circ} = 1$ so that $\|\boldsymbol{x}\|_K = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$. In other words, there is always a $\boldsymbol{y}$ so that the Generalized Cauchy-Schwarz inequality is tight. Sometimes in proofs one aims at upper bounding $\|\boldsymbol{x}\|_K$, then it can be helpful to instead upperbound $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ with $\|\boldsymbol{y}\|_{K^\circ} = 1$. We call $\boldsymbol{y}$ the *dual element* to $\boldsymbol{x}$ w.r.t. norm $\|\cdot\|_K$. Geometrically, this dual element $\boldsymbol{y}$ is the point in K° that maximizes the inner product with $\boldsymbol{x}$.



Lemma 1.7 (Existence of Dual Element)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. For any $\boldsymbol{x} \in \mathbb{R}^n$ there is a $\boldsymbol{y} \in \mathbb{R}^n$ with $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| = \|\boldsymbol{x}\|_K \cdot \|\boldsymbol{y}\|_{K^\circ}$ and $\|\boldsymbol{y}\|_{K^\circ} = 1$.*

Proof. We know by Theorem 1.4 that $\|\boldsymbol{x}\|_K = h_{K^\circ}(\boldsymbol{x}) = \max\{\langle \boldsymbol{x}, \boldsymbol{y}\rangle : \boldsymbol{y} \in K^\circ\}$. The $\boldsymbol{y}$ attaining this does the job. □

The following fact is also useful:

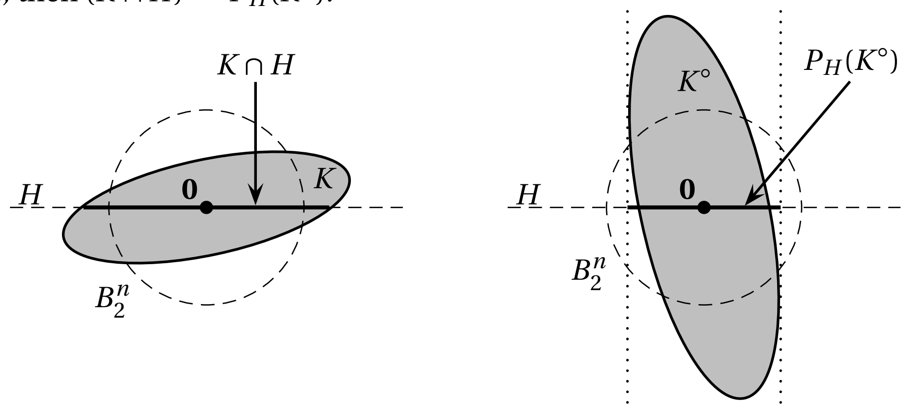
Lemma 1.8. *Let $K, Q \subseteq \mathbb{R}^n$ be convex bodies with $\mathbf{0} \in \mathrm{int}(K)$. Then $(\mathrm{conv}(K \cup Q))^\circ = K^\circ \cap Q^\circ$.*

Proof. Apply Lemma 1.3.(a). □

While most of the time we will use polarity for full dimensional convex sets, we will have one application in Chapter 7 where the sets are lower dimensional. For a subspace $H \subseteq \mathbb{R}^n$ we denote the *orthogonal projection into H* by $P_H : \mathbb{R}^n \to$

$\mathbb{R}^n$ which is the unique linear map with $P_H(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{x}$ for $\boldsymbol{x} \in H$ and $\boldsymbol{y} \in H^{\perp}$. We will now see that projection and intersection are operations that are polar to each other.

Lemma 1.9. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $H \subseteq \mathbb{R}^n$ be a subspace, then $(K \cap H)^\circ = P_H(K^\circ)$.*



Proof. First note that both objects $(K \cap H)^\circ$ and $P_H(K^\circ)$ are both contained in the subspace H and $\mathbf{0}$ is contained in the relative interior. It suffices to verify that the *support functions* for directions $\boldsymbol{y} \in H$ are identical. First we see that

$$h_{P_H(K^\circ)}(\boldsymbol{y}) \stackrel{\boldsymbol{y}\in H}{=} h_{K^\circ}(\boldsymbol{y}) \stackrel{\text{Thm } 1.4}{=} \|\boldsymbol{y}\|_K$$

as maximizing projection gives the same as maximizing over the original body K° (this uses $\boldsymbol{y} \in H$). Next, we have

$$h_{(K\cap H)^\circ}(\boldsymbol{y}) \stackrel{\text{Thm } 1.4}{=} \|\boldsymbol{y}\|_{K\cap H} \stackrel{\boldsymbol{y}\in H}{=} \|\boldsymbol{y}\|_K$$

and the claim follows. $\square$

For a matrix $\boldsymbol{A}$ and a set K, by a slight abuse of notation we write $\boldsymbol{A}(K) = \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in K\}$ as the image of K under the linear map $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$. Then it will be useful to understand how the polar of K changes if we apply a linear transformation to K:

Lemma 1.10. *Let $K \subseteq \mathbb{R}^n$ be a convex body with $\mathbf{0} \in int(K)$ and let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a regular matrix. Then $\boldsymbol{A}(K)^\circ = (\boldsymbol{A}^T)^{-1}(K^\circ) = (\boldsymbol{A}^{-1})^T(K^\circ)$.*

Proof. We have

$$\begin{aligned}
\boldsymbol{A}(K)^\circ &\stackrel{\text{Def}\circ}{=} \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{x}\rangle \le 1 \;\; \forall \boldsymbol{x} \in \boldsymbol{A}(K)\right\} \\
&= \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{A}\boldsymbol{x}\rangle \le 1 \;\; \forall \boldsymbol{x} \in K\right\} \\
&= \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{A}^T\boldsymbol{y}, \boldsymbol{x}\rangle \le 1 \;\; \forall \boldsymbol{x} \in K\right\} \\
&= \left\{(\boldsymbol{A}^T)^{-1}\boldsymbol{y} \mid \langle \boldsymbol{y}, \boldsymbol{x}\rangle \le 1 \;\; \forall \boldsymbol{x} \in K\right\} \stackrel{\text{Def}\circ}{=} (\boldsymbol{A}^T)^{-1}(K^\circ)
\end{aligned}$$

The 2nd equation holds because $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$. □

1.3 Distance measures for convex bodies

Loosely speaking, a *distance measure* for convex bodies is a quantity $d(K,Q)$ that is larger the more different the convex bodies K and Q are. But there are several meaningful distance measures and which is the "right one" will be application-dependent.

The Hausdorff Distance

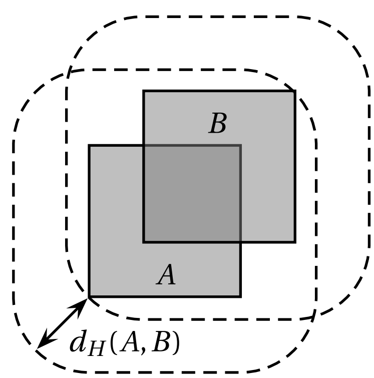
We want to discuss several distance measures on convex bodies.

Definition 1.11. We define the *Hausdorff distance* of two convex bodies $K, Q \subseteq \mathbb{R}^n$ as

$$\begin{aligned} d_H(K,Q) &:= \inf\{\delta \geq 0 \mid K \subseteq Q + \delta B_2^n \text{ and } Q \subseteq K + \delta B_2^n\} \\ &= \sup\{|h_K(\boldsymbol{u}) - h_Q(\boldsymbol{u})| \mid \boldsymbol{u} \in S^{n-1}\} \end{aligned}$$

Intuitively, d_H is the minimum radius of a ball by which one has to enlarge K and Q to include each other. Clearly $d_H(K,Q) \geq 0$ for all bodies K,Q. It is not difficult to check that:

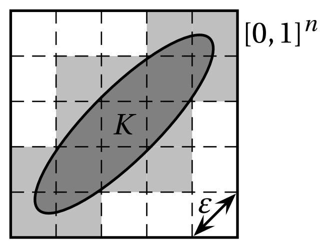
Lemma 1.12. *The Hausdorff distance is a metric and in particular $d_H(A,C) \leq d_H(A,B) + d_H(B,C)$ for convex bodies $A,B,C \subseteq \mathbb{R}^n$.*



Finally we have the following useful *compactness result*:

Theorem 1.13 (Blaschke Selection Theorem)**.** *A sequence $\{K_j\}_{j\in\mathbb{N}}$ of convex bodies with $K_j \subseteq rB_2^n$ for some fixed r has a subsequence that is convergent in the Hausdorff metric.*

Proof sketch. W.l.o.g. suppose that $K_j \subseteq [0,1]^n$ for all j. Partition the cube into an equally-spaced grid so that each cube in the grid has diameter at most $\varepsilon > 0$. Then $(\sqrt{n}/\varepsilon)^n$ many cubes suffice, but we only need that this is a finite number. Two convex bodies K_i, K_j that intersect the same set of cells have a Haussdorf distance of at most ε. Then from an infinite sequence one can iteratively filter subsequences of bodies whose distance is getting shorter and shorter quite similar to the proof of the Bolzano-Weierstrass Theorem. ☐



The geometric distance

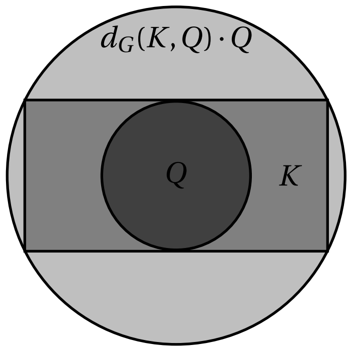
The second distance measure that we discuss will allow some transformations to the convex bodies. Formally, the *geometric distance*

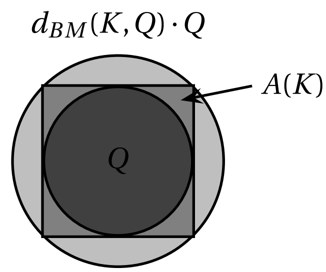
$$d_G(K,Q) := \min\left\{a \cdot b \mid \exists \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n : \frac{1}{b}(Q+\boldsymbol{y}) \subseteq K + \boldsymbol{x} \subseteq a \cdot (Q+\boldsymbol{y})\right\}$$

is the minimum factor $s \geq 1$ so that after translating and scaling with a scalar one has $Q \subseteq K \subseteq sQ$. In particular the relative position in space does not matter for this distance. This is a multiplicative distance measure with $d_G(K,Q) \geq 1$ for all K, Q.

The Banach Mazur Distance

The *Banach-Mazur distance* is defined as $d_{BM}(K,Q) := \min\{d_G(A(K),Q) \mid A \text{ linear map}\}$. Phrased differently the Banach-Mazur distance is the minimum number $s \geq 1$ so that $Q \subseteq A(K) \subseteq sQ$ where $A : \mathbb{R}^n \to \mathbb{R}^n$ is an affine map. In Chapter 2 we will see that by John's Theorem indeed $d_{BM}(K, B_2^n) \leq n$ for any convex body K. As before, the Banach-Mazur distance is a multiplicative measure with $d_{BM}(K,Q) \geq 1$ for all K, Q. We can visualize the difference between the geometric distance and the Banach-Mazur distance as follows:





geometric distance d_G Banach-Mazur distance d_{BM}

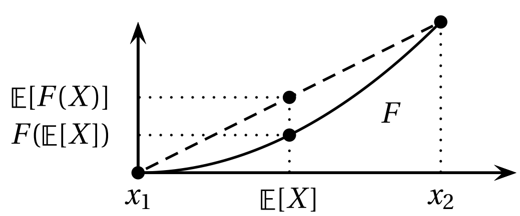
In most cases we are interested in the Banach-Mazur distance of a body $K \subseteq \mathbb{R}^n$ to the Euclidean ball. Hence we abbreviate $d_{BM}(K) := d_{BM}(K, B_2^n)$. If $F \subseteq \mathbb{R}^n$ is a subspace of dimension $k := \dim(F)$, then $K \cap F$ is a k-dimensional object. In this case, we define $d_{BM}(K \cap F) := d_{BM}(K \cap F, B_2^k)$.

1.4 Useful Inequalities

In this section, we recall several inequalities that are particularly useful when dealing with convex functions. We begin with one of the "work horses" in the area:

Theorem 1.14 (Jensen Inequality for Convex Functions)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable and $F : \mathbb{R} \to \mathbb{R}$ be a convex function. Then $F(\mathbb{E}[X]) \leq \mathbb{E}[F(X)]$.*

The inequality follows immediately from the definition of convexity.



Example of convex function F and
distribution X over only two values x_1, x_2

If the function F is rather concave then convex, then the inequality holds with reversed relation:

Theorem 1.15 (Jensen Inequality for Concave Functions)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable and $F : \mathbb{R} \to \mathbb{R}$ be a concave function. Then $F(\mathbb{E}[X]) \geq \mathbb{E}[F(X)]$.*

The next inequality is due to Young:

Theorem 1.16 (Young's Inequality)**.** *For* $x, y \geq 0$ *and* $0 \leq \lambda \leq 1$ *one has*

$$x \cdot y \leq (1-\lambda) \cdot x^{1/(1-\lambda)} + \lambda \cdot y^{1/\lambda}$$

Proof. Simply note that

$$\begin{aligned}\ln\Big((1-\lambda)x^{1/(1-\lambda)} + \lambda y^{1/\lambda}\Big) &\overset{\text{Jensen+concavity of ln}}{\geq} (1-\lambda)\ln(x^{1/(1-\lambda)}) + \lambda \ln(y^{1/\lambda}) \\ &= \ln(x) + \ln(y) = \ln(x \cdot y)\end{aligned}$$

□

Another useful inequality is the AMGM inequality:

Theorem 1.17 (Arithmetic Mean vs Geometric Mean)**.** *Let* $\alpha_1, \ldots, \alpha_n \geq 0$ *and* $x_1, \ldots, x_n \geq 0$ *and abbreviate* $\beta := \sum_{i=1}^n \alpha_i$. *Then*

$$\frac{\alpha_1 x_1 + \ldots + \alpha_n x_n}{\beta} \geq \sqrt[\beta]{x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}}$$

Proof. Let X be the random variable with $\Pr[X = x_i] = \frac{\alpha_i}{\beta}$. Then

$$\ln\Big(\sum_{i=1}^n \frac{\alpha_i}{\beta} x_i\Big) = \ln(\mathbb{E}[X]) \overset{\text{Jensen}}{\geq} \mathbb{E}[\ln(X)] = \sum_{i=1}^n \frac{\alpha_i}{\beta} \ln(x_i) = \ln\Big(\prod_{i=1}^n x_i^{\alpha_i/\beta}\Big)$$

as ln is concave. □

The inequality of Hölder is basically a generalization of Cauchy-Schwarz to general $\|\cdot\|_p$-norms:

Theorem 1.18 (Hölder's Inequality I)**.** *Let* $X, Y : \Omega \to \mathbb{R}_{\geq 0}$ *be jointly distributed non-negative random variables. Then for all* $0 \leq \lambda \leq 1$ *one has* $\mathbb{E}[X^{1-\lambda} Y^\lambda] \leq \mathbb{E}[X]^{1-\lambda} \mathbb{E}[Y]^\lambda$.

Proof. Scaling X by $s > 0$ scales both sides of the inequality by the same factor of $s^{1-\lambda}$. Hence we may assume w.l.o.g. that $\mathbb{E}[X] = 1$; similarly assume $\mathbb{E}[Y] = 1$. Then applying Young's Inequality gives

$$\begin{aligned}\mathbb{E}[X^{1-\lambda} Y^\lambda] &\overset{\text{Young}}{\leq} \mathbb{E}\Big[(1-\lambda) \cdot (X^{1-\lambda})^{\frac{1}{1-\lambda}} + \lambda \cdot (Y^\lambda)^{\frac{1}{\lambda}}\Big] = \mathbb{E}[(1-\lambda)X + \lambda Y] \\ &= (1-\lambda)\mathbb{E}[X] + \lambda \mathbb{E}[Y] = 1 = \mathbb{E}[X]^{1-\lambda} \cdot \mathbb{E}[Y]^\lambda\end{aligned}$$

□

Often, Hölder is stated in a different but equivalent form:

Theorem 1.19 (Hölder's Inequality II). *Let $X, Y : \Omega \to \mathbb{R}$ be jointly distributed random variables. Let $p, q \geq 1$ be a pair with $\frac{1}{p} + \frac{1}{q} = 1$. Then $\mathbb{E}[|X \cdot Y|] \leq \mathbb{E}[|X|^p]^{1/p} \cdot \mathbb{E}[|Y|^q]^{1/q}$.*

We conclude with Minkowsi's Inequality which in fact provides the proof that the bodies B_p^n are convex.

Lemma 1.20 (Minkowski's Inequality I). *Let $1 \leq p < \infty$ and let X, Y be jointly distributed random variables so that $\mathbb{E}[|X|^p], \mathbb{E}[|Y|^p] < \infty$. Then $\mathbb{E}[|X+Y|^p]^{1/p} \leq \mathbb{E}[|X|^p]^{1/p} + \mathbb{E}[|Y|^p]^{1/p}$.*

Proof. First note that $|x+y|^p \leq 2^p \cdot (|x|^p + |y|^p)$ and so $\mathbb{E}[|X+Y|^p]^{1/p} < \infty$. Now, rescale the random variables so that $\mathbb{E}[|X+Y|^p] = 1$. Choose q so that $\frac{1}{p} + \frac{1}{q} = 1 \Leftrightarrow q = \frac{p}{p-1}$. Then

$$\begin{aligned} 1 &= \mathbb{E}\left[|X+Y|^p\right] = \mathbb{E}\left[|X| \cdot |X+Y|^{p-1}\right] + \mathbb{E}\left[|Y| \cdot |X+Y|^{p-1}\right] \\ &\overset{\text{Hölder II}}{\leq} \left(\mathbb{E}\left[|X|^p\right]^{1/p} + \mathbb{E}\left[|Y|^p\right]^{1/p}\right) \cdot \Big(\underbrace{\mathbb{E}\left[|X+Y|^{q(p-1)}\right]}_{=1}\Big)^{1/q} = \mathbb{E}\left[|X|^p\right]^{1/p} + \mathbb{E}\left[|Y|^p\right]^{1/p} \end{aligned}$$

as $q(p-1) = p$. □

Lemma 1.21 (Minkowski's Inequality II). *Let $1 \leq p < \infty$, let $\|\cdot\|_K$ be a norm and let $\boldsymbol{X}, \boldsymbol{Y}$ be jointly distributed random variables on $\mathbb{R}^n$ so that $\mathbb{E}[\|\boldsymbol{X}\|_K^p], \mathbb{E}[\|\boldsymbol{Y}\|_K^p] < \infty$. Then $\mathbb{E}[\|\boldsymbol{X}+\boldsymbol{Y}\|_K^p]^{1/p} \leq \mathbb{E}[\|\boldsymbol{X}\|_K^p]^{1/p} + \mathbb{E}[\|\boldsymbol{Y}\|_K^p]^{1/p}$.*

Proof. We bound $\mathbb{E}[\|\boldsymbol{X}+\boldsymbol{Y}\|_K^p]^{1/p} \leq \mathbb{E}[|\|\boldsymbol{X}\|_K + \|\boldsymbol{Y}\|_K|^p]^{1/p} \leq \mathbb{E}[\|\boldsymbol{X}\|_K^p]^{1/p} + \mathbb{E}[\|\boldsymbol{Y}\|_K^p]^{1/p}$ using the triangle inequality for $\|\cdot\|_K$ and Minkowski's Inequality I (Lem 1.20) for the random variables $\|\boldsymbol{X}\|_K$ and $\|\boldsymbol{Y}\|_K$. □

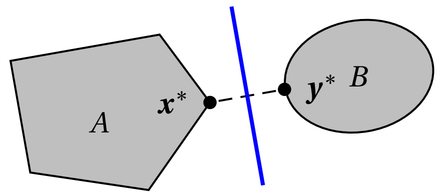
1.5 The Hahn-Banach Theorem and relatives

A simple, but powerful result is the following:

Theorem 1.22 (Separating Hyperplane Theorem I). *Let $A, B \subseteq \mathbb{R}^n$ non-empty disjoint convex sets where both A and B are closed and at least one of them is bounded. Then there is a vector $\boldsymbol{c} \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$ so that*

$$\langle \boldsymbol{c}, \boldsymbol{x} \rangle > \delta > \langle \boldsymbol{c}, \boldsymbol{y} \rangle \quad \forall \boldsymbol{x} \in A \ \forall \boldsymbol{y} \in B$$

Sketch. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in A \times B$ be the pair minimizing the distance $\|\boldsymbol{x}^* - \boldsymbol{y}^*\|_2$ (this must exist for the following reason: suppose that A is bounded; then A is compact. Then the distance function $d(\boldsymbol{x}) := \min\{\|\boldsymbol{y}-\boldsymbol{x}\|_2 \mid \boldsymbol{y} \in B\}$ is well-defined and continuous, hence a minimum is attained on P). Then the hyperplane through $\frac{1}{2}(\boldsymbol{x}^* + \boldsymbol{y}^*)$ with normal vector $\boldsymbol{c} = \boldsymbol{x}^* - \boldsymbol{y}^*$ separates A and B.



□

The statement also holds for unbounded sets — as long as one is willing to give up the strict separation:

Theorem 1.23 (Separating Hyperplane Theorem II)**.** *Let $A, B \subseteq \mathbb{R}^n$ be non-empty disjoint convex sets. Then there is a vector $\boldsymbol{c} \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$ so that*

$$\langle \boldsymbol{c}, \boldsymbol{x} \rangle \geq \delta \geq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \quad \forall \boldsymbol{x} \in A \ \forall \boldsymbol{y} \in B$$

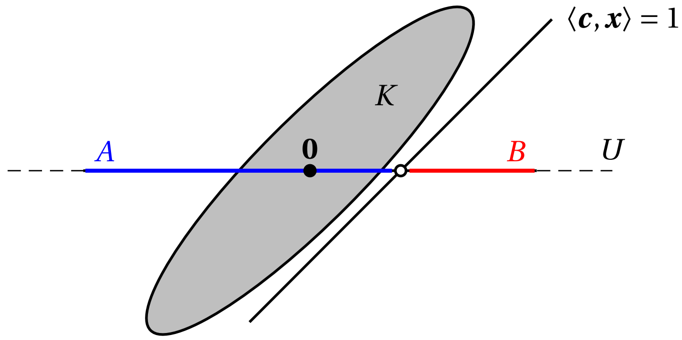
A *Banach space* is a pair $X = (V, \|\cdot\|)$, where V is an $\mathbb{R}$-vector space[1] and $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ is a norm. Most of the time the vector space is simply $V = \mathbb{R}^n$ but in convex geometry we will find other infinite-dimensional vector spaces occuring. For example $X = (V, \|\cdot\|)$ with $V := \{f : [a,b] \to \mathbb{R} \mid f \text{ continuous}\}$ and $\|f\| := \max_{x \in [a,b]} |f(x)|$ is such a vector space.

Theorem 1.24 (Hahn-Banach Theorem)**.** *Let $(V, \|\cdot\|)$ be a Banach space and let $U \subseteq V$ be a subspace. Suppose $F : U \to \mathbb{R}$ is a linear function with $F(\boldsymbol{x}) \leq \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in U$. Then there exists a linear function $\tilde{F} : V \to \mathbb{R}$ so that*

- *$\tilde{F}(\boldsymbol{x}) = F(\boldsymbol{x})$ for all $\boldsymbol{x} \in U$.*
- *$\tilde{F}(\boldsymbol{x}) \leq \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in V$.*

Sketch. We will only prove the statement for the finite-dimensional case, i.e. $V = \mathbb{R}^n$. Define $K := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \leq 1\}$ as the unit ball of the norm $\|\cdot\|$. Let $A := \{\boldsymbol{x} \in U \mid F(\boldsymbol{x}) \leq 1\}$ and $B := \{\boldsymbol{x} \in U \mid F(\boldsymbol{x}) > 1\}$. Then $\mathrm{conv}(K \cup A)$ and B are convex and disjoint, hence there exists a separating hyperplane of the form $\langle \boldsymbol{c}, \boldsymbol{x} \rangle = \delta$. Scale this one so that $\delta = 1$ (works as $\mathbf{0} \in \mathrm{int}(\mathrm{conv}(K \cup A))$). Then $\tilde{F}(\boldsymbol{x}) := \langle \boldsymbol{c}, \boldsymbol{x} \rangle$ does the job.

[1] One can also consider Banach spaces with $\mathbb{C}$-vector spaces, but for the scope of this text we restrict to to $\mathbb{R}$ as the underlying field.



□

1.6 Steiner symmetrization

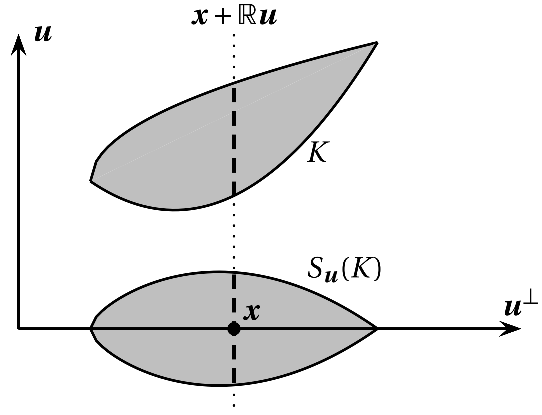
There are several cases of inequalities for convex bodies K where the worst case is attained for Euclidean balls. Then, one standard proof technique is to gradually transform K into a ball. This is done by the so-called *Steiner symmetrization*. For $\boldsymbol{u} \in \mathbb{R}^n$ we define $\boldsymbol{u}^\perp := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{u} \rangle = 0\}$ as the $(n-1)$-dimensional subspace that is orthogonal to $\boldsymbol{u}$.

Definition 1.25. Let $K \subseteq \mathbb{R}^n$ be a convex body and $\boldsymbol{u} \in S^{n-1}$ a unit direction. The *Steiner symmetral* $S_{\boldsymbol{u}}(K) \subseteq \mathbb{R}^n$ of K in direction $\boldsymbol{u}$ is defined so that for every $\boldsymbol{x} \in \boldsymbol{u}^\perp$ one has

$$\mathrm{Vol}_1((\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap K) = \mathrm{Vol}_1((\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap S_{\boldsymbol{u}}(K))$$

where $(\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap S_{\boldsymbol{u}}(K)$ is an interval centered at $\boldsymbol{x}$.

The way to interpret this construction is as follows: take a body K and a point $\boldsymbol{x} \in \boldsymbol{u}^\perp$. Then shift the interval $(\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap K$ until it is centered around $\boldsymbol{x}$. The new body is then called $S_{\boldsymbol{u}}(K)$. Note that in particular the body $S_{\boldsymbol{u}}(K)$ is symmetric w.r.t. the hyperplane $\boldsymbol{u}^\perp$.



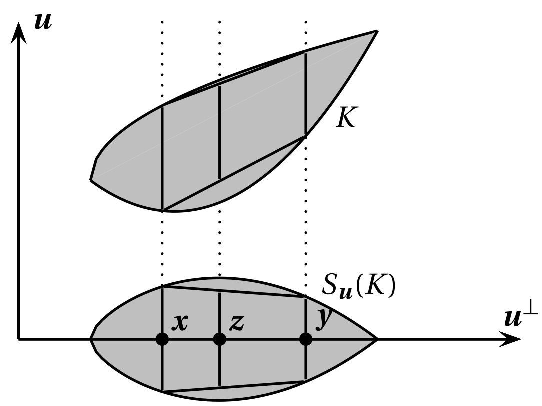
The Steiner symmetrization has many useful properties. For example it preserves volume and convexity. In particular:

Theorem 1.26 (Properties of Steiner Symmetrization)**.** *For convex bodies $K, T \subseteq \mathbb{R}^n$ and $\boldsymbol{u} \in S^{n-1}$ one has*

1. *K convex $\Rightarrow S_{\boldsymbol{u}}(K)$ convex.*
2. *$\lambda S_{\boldsymbol{u}}(K) = S_{\boldsymbol{u}}(\lambda K)$*
3. *$S_{\boldsymbol{u}}(K) + S_{\boldsymbol{u}}(T) \subseteq S_{\boldsymbol{u}}(K+T)$*
4. *$S_{\boldsymbol{u}}$ is continuous w.r.t. Haussdorf distance.*
5. *$\mathrm{Vol}_n(S_{\boldsymbol{u}}(K)) = \mathrm{Vol}_n(K)$.*
6. *$\partial(S_{\boldsymbol{u}}(K)) \leq \partial(K)$, where $\partial(K)$ denotes the surface area of K.*
7. *$\mathrm{diam}(S_{\boldsymbol{u}}(K)) \leq \mathrm{diam}(K)$.*
8. *$\mathrm{inradius}(S_{\boldsymbol{u}}(K)) \geq \mathrm{inradius}(K)$ and $\mathrm{circumrad}(S_{\boldsymbol{u}}(K)) \leq \mathrm{circumrad}(K)$.*

Here inradius(K) is the largest r so that $\boldsymbol{c} + rB_2^n \subseteq K$ for some center $\boldsymbol{c}$. Moreover, circumradius(K) is the minimum radius r so that there is a $\boldsymbol{c}$ with $K \subseteq \boldsymbol{c} + rB_2^n$.

To see convexity of the Steiner symmetral, we need to argue that the line segment between two points in $S_{\boldsymbol{u}}(K)$ is again included in $S_{\boldsymbol{u}}(K)$. We can write the two candidate points as $\boldsymbol{x} + s \cdot \boldsymbol{u}$ and $\boldsymbol{y} + t \cdot \boldsymbol{u}$ with $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{u}^{\perp}$ and $s, t \in \mathbb{R}$. Let $\ell_{\boldsymbol{x}} := \mathrm{Vol}_1(K \cap (\boldsymbol{x} + \mathbb{R}\boldsymbol{u}))$ be the length of the interval of K intersected with the line $\boldsymbol{x} + \mathbb{R}\boldsymbol{u}$. Then it suffices to check that for $0 < \lambda < 1$ and $\boldsymbol{z} = (1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y}$ one has $\ell_{\boldsymbol{z}} \geq (1-\lambda)\ell_{\boldsymbol{x}} + \lambda \ell_{\boldsymbol{y}}$. And indeed this follows from the convexity of K itself.



The usefulness of the Steiner symmetrization is that we can use it to transform every convex body into a ball. Formally, one can prove:

Theorem 1.27. *Let $K \subseteq \mathbb{R}^n$ be a convex body and let r be the radius so that $Vol_n(K) = Vol_n(rB_2^n)$. Then there exists a sequence of vectors $\boldsymbol{u}_j \in S^{n-1}$ so that the sequence $K_j := S_{\boldsymbol{u}_j}(K_{j-1})$ with $K_0 := K$ converges to rB_2^n w.r.t. the Hausdorff metric.*

1.6.1 Urysohn's Inequality

We will now see a quick application of the Steiner symmetrization to prove *Urysohn's Inequality*. Recall that for a convex body K and a vector $\boldsymbol{u} \in \mathbb{R}^n$ with $\|\boldsymbol{u}\|_2 = 1$, the *support function* is $h_K(\boldsymbol{u}) = \max\{\langle \boldsymbol{u}, \boldsymbol{x}\rangle \mid \boldsymbol{x} \in K\}$ and $w_K(\boldsymbol{u}) = h_K(\boldsymbol{u}) + h_K(-\boldsymbol{u})$ is the (geometric) *width* in direction $\boldsymbol{u}$. Then the *mean width* is simply $w(K) := \mathbb{E}_{\boldsymbol{u}\sim S^{n-1}}[w_K(\boldsymbol{u})]$, where $\boldsymbol{u} \sim S^{n-1}$ picks a uniform random unit vector. Now we will see the very intuitive fact that among all convex bodies of the same volume, the ball minimizes the mean width.

Theorem 1.28 (Urysohn)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then*

$$w(K) \geq 2 \cdot \Big(\frac{Vol_n(K)}{Vol_n(B_2^n)}\Big)^{1/n}$$

Suppose we scale K so that $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(B_2^n)$. We could apply Steiner Symmetrization until the body converges to B_2^n — all we need to show is that the mean width is not increasing:

Lemma 1.29. *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then $w(S_{\boldsymbol{\theta}}(K)) \leq w(K)$.*

Proof. After rotation we may assume that $\boldsymbol{\theta} = \boldsymbol{e}_n$ meaning that the symmetrization happens for the last coordinate. We will write $(\boldsymbol{x}, t) \in K$ where $\boldsymbol{x} \in \mathbb{R}^{n-1}$ and $t \in \mathbb{R}$. Observe that $((\boldsymbol{x}, t_1)$ and $(\boldsymbol{x}, t_2) \in K) \Leftrightarrow (\boldsymbol{x}, \frac{t_1-t_2}{2}) \in S_{\boldsymbol{\theta}}(K)$. Then the support function of the Steiner symmetrization is

$$\begin{aligned}
h_{S_{\boldsymbol{\theta}}(K)}(\boldsymbol{u}) &= \max\Big\{\langle(\boldsymbol{x}, \frac{t_1-t_2}{2}), \boldsymbol{u})\rangle : (\boldsymbol{x}, t_1) \in K \text{ and } (\boldsymbol{x}, t_2) \in K\Big\} \\
&\leq \frac{1}{2}\cdot\Big(\max\{\langle(\boldsymbol{x}, t_1), \boldsymbol{u}\rangle : (\boldsymbol{x}, t_1) \in K\} + \max\{\langle(\boldsymbol{x}, -t_2), \boldsymbol{u}\rangle : (\boldsymbol{x}, t_2) \in K\}\Big) \\
&= \frac{1}{2}\Big(h_K(\boldsymbol{u}) + h_K(\boldsymbol{u}')\Big)
\end{aligned}$$

where we write $\boldsymbol{u}' := (u_1, \ldots, u_{n-1}, -u_n)$ as the vector $\boldsymbol{u}$ with flipped last coordinate. Note that for $\boldsymbol{u} \sim S^{n-1}$ the expectation of $h_K(\boldsymbol{u})$ and $h_K(\boldsymbol{u}')$ is identical. Hence $\mathbb{E}_{\boldsymbol{u}\sim S^{n-1}}[h_{S_{\boldsymbol{\theta}}(K)}(\boldsymbol{u})] \leq \mathbb{E}_{\boldsymbol{u}\sim S^{n-1}}[h_K(\boldsymbol{u})]$ and the claim follows. $\square$

1.6.2 Blaschke-Santaló Inequality

The Blaschke-Santaló inequality is another example where some quantity is maximized or mininimized for Euclidean balls.

Theorem 1.30. *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex body. Then*

$$Vol_n(K) \cdot Vol_n(K^\circ) \leq Vol_n(B_2^n)^2.$$

The proof works again using Steiner symmetrization. By construction we have $\mathrm{Vol}_n(S_{\boldsymbol{\theta}}(K)) = \mathrm{Vol}_n(K)$ for any direction $\boldsymbol{\theta} \in S^{n-1}$. The non-trivial part is to prove that the volume of the polar does not decrease in one symmetrization step:

Lemma 1.31. *For any symmetric convex body $K \subseteq \mathbb{R}^n$ and any $\boldsymbol{\theta} \in S^{n-1}$ one has $Vol_n(K^\circ) \leq Vol_n((S_{\boldsymbol{\theta}}(K))^\circ)$.*

For a symmetric convex body $K \subseteq \mathbb{R}^n$, we define the *Mahler product* as $s(K) := \mathrm{Vol}(K) \cdot \mathrm{Vol}_n(K^\circ)$. With this lemma we know that $s(K) \leq s(S_{\boldsymbol{\theta}}(K))$ and hence by a limit argument we can derive that $s(K) \leq s(B_2^n)$, which settles the claim. The proof of Lemma 1.31 is not too hard but as we will see a structurally stronger statement in Chapter 8 we skip it here.

1.7 Brunn's Concavity Principle and Log-Concavity

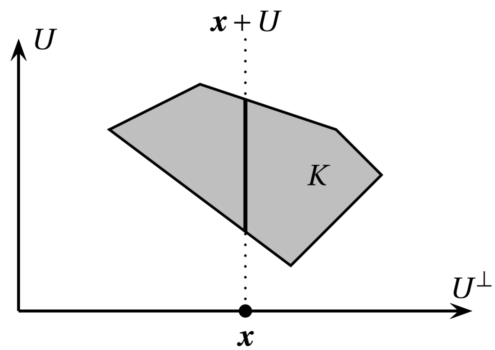
In this section, we will discuss how the volume of *slices* of convex bodies behave. If $U \subseteq \mathbb{R}^n$ is a subspace with dimension $k = \dim(U)$ and $K \subseteq \mathbb{R}^n$ is some set then we denote $\mathrm{Vol}_k(K \cap U)$ as the k-dimensional volume of $K \cap U$ that "lives" inside the subspace U. More formally one could define this quantity by picking any orthonormal basis $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ for U and setting $\mathrm{Vol}_k(K \cap U) := \mathrm{Vol}_k(\{\boldsymbol{y} \in \mathbb{R}^k \mid \sum_{i=1}^k y_i \boldsymbol{u}_i \in K\})$. Analogously, if U is an affine subspace.

For a subspace $U \subseteq \mathbb{R}^n$ we write $U^\perp := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \perp \boldsymbol{y} \ \forall \boldsymbol{y} \in U\}$ as the $(n - \dim(U))$-dimensional subspace that is orthogonal to U.

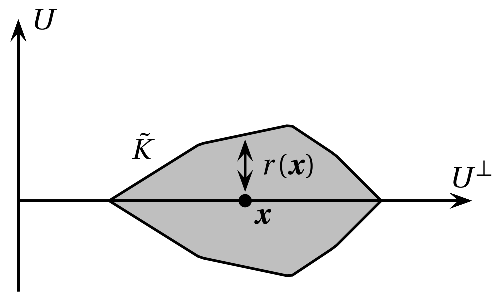
Theorem 1.32 (Brunn's Concavity Principle I)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body and let $U \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Then the function $F : U^\perp \to \mathbb{R}$ defined by*

$$F(\boldsymbol{x}) := Vol_k(K \cap (U + \boldsymbol{x}))^{1/k}$$

is concave on its support.



Proof. We apply the Steiner Symmetrization for directions $\boldsymbol{u} \in U$ and obtain a limiting convex body $\tilde{K}$ so that $\tilde{K} \cap (U + \boldsymbol{x})$ is a k-dimensional ball of some radius $r(\boldsymbol{x})$ and the volumes of intersections with translates of U have not changed, meaning that $\mathrm{Vol}_k(\tilde{K} \cap (U + \boldsymbol{x})) = \mathrm{Vol}_k(K \cap (U + \boldsymbol{x}))$ for all $\boldsymbol{x} \in U^\perp$.



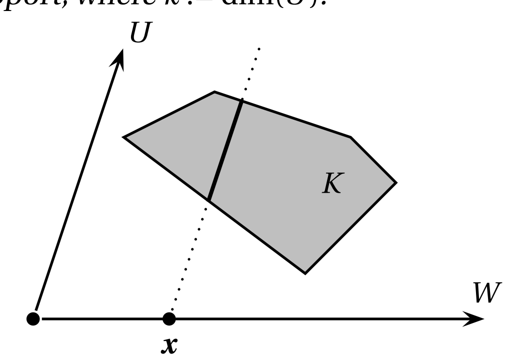
Then by convexity of $\tilde{K}$, $r(\boldsymbol{x})$ is concave and so is F. $\square$

It will be useful to make the observation that the subspace that form the domain of F and the subspace used for slicing do not need to be orthogonal.

Theorem 1.33 (Brunn's Concavity Principle II)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body and let $U, W \subseteq \mathbb{R}^n$ be subspace. Then the function $F : W \to \mathbb{R}$ defined by*

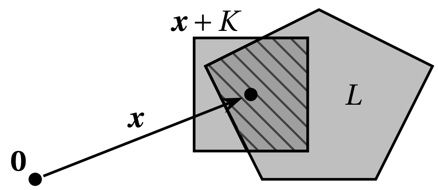
$$F(\boldsymbol{x}) := Vol_k(K \cap (U + \boldsymbol{x}))^{1/k}$$

is concave on its support, where $k := \dim(U)$.



A useful inequality for later will be that also the intersection of convex sets behaves in a log-concave manner:

Lemma 1.34. *Let $K, L \subseteq \mathbb{R}^n$ be convex sets. Then the function $F : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $F(\boldsymbol{x}) := Vol_n((\boldsymbol{x}+K) \cap L)^{1/n}$ is concave.*



Proof. We rewrite the function F as

$$\begin{aligned} F(\boldsymbol{x}) &= \mathrm{Vol}_n\left(\{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y}-\boldsymbol{x} \in K \text{ and } \boldsymbol{y} \in L\}\right)^{1/n} \\ &= \frac{1}{\sqrt{2}}\mathrm{Vol}_n\left(\{(\boldsymbol{y}-\boldsymbol{x},\boldsymbol{y}) \in K \times L\}\right)^{1/n} = \frac{1}{\sqrt{2}}\mathrm{Vol}_n\left(\{(K \times L) \cap (U + (-\boldsymbol{x},\mathbf{0}))\}\right)^{1/n} \end{aligned}$$

where we define a subspace $U := \{(\boldsymbol{y},\boldsymbol{y}) \mid \boldsymbol{y} \in \mathbb{R}^n\}$ with $\dim(U) = n$. By Brunn's Concavity principle, such a function is concave on its support. $\square$

A function $F : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is called *log concave* if $F(\boldsymbol{x}) = \exp(-G(\boldsymbol{x}))$ for some convex function $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Equivalently F is log-concave if $\ln(F(\boldsymbol{x}))$ is concave. A third definition is that

$$F(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq F(\boldsymbol{x})^{\lambda} \cdot F(\boldsymbol{y})^{1-\lambda}$$

for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$. Log concave functions arise very naturally in the context of convex geometry. For example, the Gaussian density function $\gamma_n(\boldsymbol{x})$ is log-concave. Log-concavity is actually a weaker property than concavity:

Lemma 1.35. *For each function $F : \mathbb{R}^n \to \mathbb{R}_{>0}$ one has: F concave $\Rightarrow$ F log-concave.*

Proof. It suffices to verify this for $n = 1$ as the definitions are properties for lines. Suppose that F is concave and w.l.o.g. suppose that F is differentiable. Then $F''(x) \leq 0$ for all x. Hence $(\ln F(x))'' = \frac{F''(x)}{F(x)} - \frac{F'(x)^2}{F(x)^2} \leq 0$ as well. $\square$

For example, we have proven that the function $\mathrm{Vol}_n((\boldsymbol{x}+K) \cap L)^{1/n}$ is concave, and so $\ln(\mathrm{Vol}_n((\boldsymbol{x}+K) \cap L)^{1/n}) = \frac{1}{n}\ln(\mathrm{Vol}_n((\boldsymbol{x}+K) \cap L))$ is concave. That means the function $\mathrm{Vol}_n((\boldsymbol{x}+K) \cap L)$ is log-concave without the need for the exponent $1/n$.

Corollary 1.36. Let K be a convex set, $\boldsymbol{\theta} \in \mathbb{R}^n$ and define a function $G : \mathbb{R} \to \mathbb{R}_{\geq 0}$ by $G(t) := \mathrm{Vol}_n(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \leq t\})$. Then $G(t)^{1/n}$ is concave on its support and $G(t)$ itself is log concave on its support.

Proof. W.l.o.g. suppose $\|\boldsymbol{\theta}\|_2 = 1$. If we set $F(\boldsymbol{x}) := \mathrm{Vol}_n\big(K \cap (\boldsymbol{x} + \{\boldsymbol{y} \in \mathbb{R}^n : \langle \boldsymbol{y}, \boldsymbol{\theta} \rangle \leq 0\})\big)$ then by Lemma 1.34, $F(\boldsymbol{x})^{1/n}$ is concave on its support. Then the same holds for $G(t) = F(t\boldsymbol{\theta})$ which is the restriction to a line. □

We want to spent a few words on the behavior of log concave functions. For the sake of simplicity, consider a 1-dimensional log-concave function $G : \mathbb{R} \to \mathbb{R}_{>0}$. Then G needs not to be concave. But $\ln G(t)$ is concave. Hence for any fixed point $t^* \in \mathbb{R}$ one has

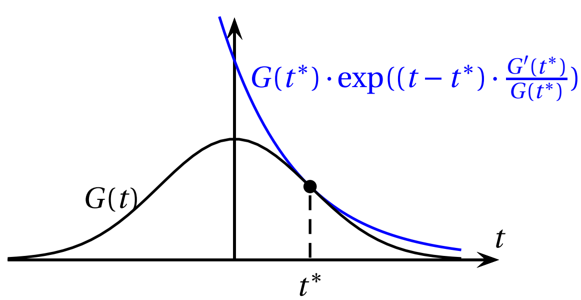
$$\ln(G(t)) \leq \ln(G(t^*)) + (t - t^*) \cdot \ln(G(t^*))' = \ln(G(t^*)) + (t - t^*) \cdot \frac{G'(t^*)}{G(t^*)} \quad \forall t \in \mathbb{R}$$

Now exponentiating this inequality gives the following useful estimate:

Lemma 1.37. *Let $G : \mathbb{R} \to \mathbb{R}_{>0}$ be log concave. Then for $t^* \in \mathbb{R}$ one has*

$$G(t) \leq G(t^*) \cdot \exp\left((t - t^*) \cdot \frac{G'(t^*)}{G(t^*)}\right)$$

In particular if we have any point t^* with $G'(t^*) < 0$, then the log concave function must be decaying at least at an exponential rate.



Example: $G(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$

Grünbaum's Lemma

We want to show an application of Brunn's Concavity Principle to prove a beautiful lemma by Grünbaum:

> *Any hyperplane through the barycenter of a convex body $K \subseteq \mathbb{R}^n$ has at least a $\frac{1}{e}$ fraction of the volume on each side.*

Recall that the *barycenter* of a set K is $\mathbb{E}_{\boldsymbol{x}\sim K}[\boldsymbol{x}] = \frac{1}{\mathrm{Vol}_n(K)}\int_K \boldsymbol{x}\, d\boldsymbol{x}$. In particular the barycenter is $\mathbf{0}$ if and only if $\int_{\mathbb{R}} t \cdot \mathrm{Vol}_{n-1}(\{\boldsymbol{x} \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle = t\})dt = 0$ for every direction $\boldsymbol{\theta} \in S^{n-1}$.

Lemma 1.38 (Grünbaum)**.** *Let K be a convex set with $Vol_n(K) = 1$ and the barycenter at $\mathbf{0}$. Then for every $\boldsymbol{\theta} \in S^{n-1}$ one has $Vol_n(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle \leq 0\}) \geq \frac{1}{e}$.*

Proof. After scaling we may assume that $\boldsymbol{x} \in K \Rightarrow -1 \leq \langle \boldsymbol{\theta}, \boldsymbol{x}\rangle \leq 1$. Consider the function

$$G(t) := \mathrm{Vol}_n(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle \leq t\})$$

Then by assumption $G(-1) = 0$ and $G(1) = 1$. It is not hard to see that the derivative of that function is $G'(t) = \mathrm{Vol}_{n-1}(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle = t\})$. Next, we will argue that the graph of the function G partitions the box $[-1,1]\times[0,1]$ into two parts of equal area. Note that this is due to $\mathbf{0}$ being the barycenter.

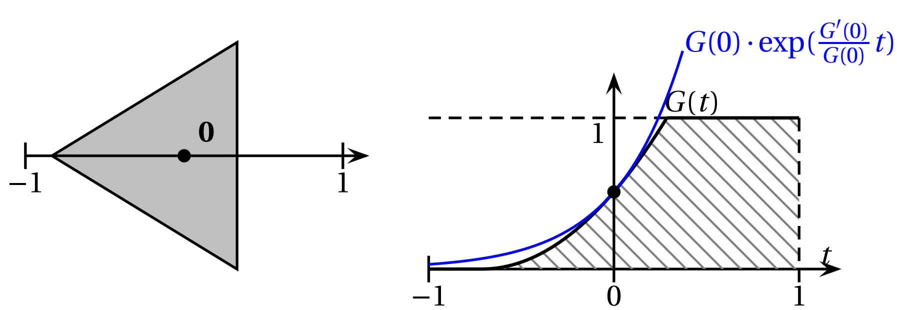
Claim. *One has $\int_{-1}^{1} G(t)dt = 1$.*

Proof of claim. As $\mathbf{0}$ is the barycenter we have

$$0 \overset{\text{barycenter}}{=} \int_{-1}^{1} t \cdot G'(t)dt \overset{\text{integration by parts}}{=} \underbrace{[t \cdot G(t)]_{t=-1..1}}_{=1} - \int_{-1}^{1} G(t) \cdot 1\, dt$$

Rearranging gives the claim. $\square$

It remains to show that $G(0) \geq \frac{1}{e}$. The intuition for the proof is that if $G(0)$ was too small, then the area below the curve could not be as large as 1. By Cor. 1.36, we know that the function $G(t)$ is log-concave. We can apply the upper bound for log-concave functions from Lemma 1.37 for $t^* = 0$ to obtain that $G(t) \leq G(0) \cdot \exp\left(\frac{G'(0)}{G(0)} t\right)$ for all t.



Then we upper bound the area below the curve G as

$$\begin{aligned} 1 &= \int_{-1}^{1} G(t)\,dt \leq \int_{-1}^{1} \min\Big\{\exp\Big(\frac{G'(0)}{G(0)}t\Big),1\Big\}dt = G(0)\int_{-1}^{\alpha}\exp\Big(\frac{G'(0)}{G(0)}t\Big)dt + \underbrace{\int_{\alpha}^{1} 1\,dt}_{=1-\alpha} \\ &\leq G(0)\cdot\Big[\frac{G(0)}{G'(0)}\exp\Big(\frac{G'(0)}{G(0)}t\Big)\Big]_{t=-\infty}^{\alpha} + (1-\alpha) \\ &= \frac{G(0)^2}{G'(0)}\exp\Big(\frac{G'(0)}{G(0)}\cdot\alpha\Big) + (1-\alpha) \stackrel{\alpha:=\frac{G(0)}{G'(0)}}{=} 1 + (e\cdot G(0)-1)\cdot\frac{G(0)}{G'(0)} \end{aligned}$$

Rearranging gives the desired claim of $G(0) \geq \frac{1}{e}$. □

1.8 The Brunn-Minkowski inequality

One of the most often used inequalities in convex geometry is the *Brunn-Minkowski Inequality* which gives lower bounds on the volume of the Minkowski sum. In the most simple form it is as follows:

> *For any two sets $A, B \subseteq \mathbb{R}^n$ with $Vol_n(A) = 1 = Vol_n(B)$ and $0 \leq \lambda \leq 1$ one has $Vol_n(\lambda A + (1-\lambda)B) \geq 1$.*

It is remarkable that the inequality makes no restriction to the shape of A and B. Note that, for example if $A = B$, then the inequality is tight. On the other hand, for sets that have a very different shape the volume of the Minkowski sum might be a lot larger.

If A and B are not of identical volume, one needs to find the right normalization. Formally one can state:

Theorem 1.39 (Brunn-Minkowski Inequality I)**.** *Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then for $0 < \lambda < 1$ one has*

$$Vol_n(\lambda A + (1-\lambda)B)^{1/n} \geq \lambda \cdot Vol_n(A)^{1/n} + (1-\lambda)\cdot Vol_n(B)^{1/n}.$$

One can also rewrite the inequality as

Theorem 1.40 (Brunn-Minkowski Inequality II)**.** *Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then*

$$Vol_n(A+B)^{1/n} \geq Vol_n(A)^{1/n} + Vol_n(B)^{1/n}.$$

Another more multiplicative form is:

Theorem 1.41 (Brunn-Minkowski Inequality III)**.** *Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then for $0 < \lambda < 1$ one has*

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

There is also a "fractional" or "functional" version of the Brunn-Minkowski Inequality which is as follows:

Theorem 1.42 (Prékopa-Leindler Inequality)**.** *For $0 < \lambda < 1$, let $f, g, h : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be measurable functions so that*

$$h(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq f(\boldsymbol{x})^{\lambda} g(\boldsymbol{y})^{1-\lambda} \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Then

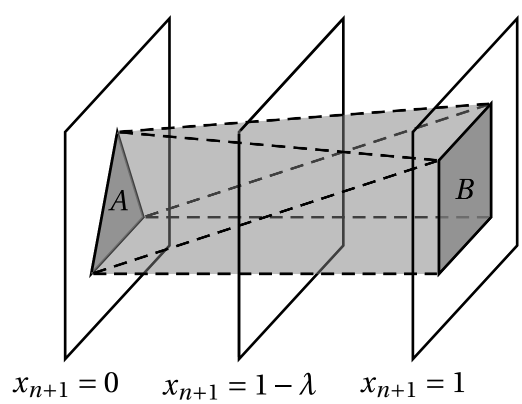
$$\int_{\mathbb{R}^n} h(\boldsymbol{x}) d\boldsymbol{x} \geq \Big(\int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x} \Big)^{\lambda} \cdot \Big(\int_{\mathbb{R}^n} g(\boldsymbol{x}) d\boldsymbol{x} \Big)^{1-\lambda}$$

It might be worth noting that if the premise holds for a particular value of λ (even if not for all), then also the conclusion holds for that same value. One can observe that the Prékopa-Leindler inequality is very similar to the multiplicative version of Brunn-Minkowski (BM 3).

There is a simple proof for Brunn-Minkowski for the case that the sets A, B are convex. Also historically, this was the first approach.

Theorem 1.43. *Brunn-Minkowski Inequality I holds for convex sets $A, B \in \mathbb{R}^n$.*

Proof. We embed the sets A and B in the two parallel planes $x_{n+1} = 0$ and $x_{n+1} = 1$ of $\mathbb{R}^{n+1}$, resp. Then we consider $K := \text{conv}((A \times \{0\}) \cup (B \times \{1\}))$ and the slices $K(t) := \{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x}, t) \in K\}$.



Observe that $K(0) = A$ and $K(1) = B$ and more generally $K(t) = (1-t) \cdot A + t \cdot B$. We know from Brunn's Concavity Principle that functions of the form $t \to \text{Vol}_n(K(t))$

are log-concave. Then

$$\mathrm{Vol}_n(\underbrace{K(\lambda\cdot 0+(1-\lambda)\cdot 1)}_{=\lambda A+(1-\lambda)B}) \geq \mathrm{Vol}_n(\underbrace{K(0)}_{=A})^{\lambda}\cdot \mathrm{Vol}_n(\underbrace{K(1)}_{=B})^{1-\lambda}$$

and the claim follows. $\square$

1.8.1 A proof of Brunn-Minkowski inequality III and Prékopa-Leindler Inequality

We will now prove the Brunn-Minkowski inequality III and Prékopa-Leindler Inequality together in several steps. Here we are taking a part of the proof from Ball's survey [Bal97].

Claim I. *Brunn-Minkowski Inequality I+III holds for $n = 1$.*

Proof of claim. Take compact subsets $A, B \subseteq \mathbb{R}$. Translate the sets so that $\max\{x \in A\} = 0 = \min\{x \in B\}$ and abbreviate $C := \lambda A + (1-\lambda)B$. Since $0 \in A\cap B$, we have $\lambda A \subseteq C$ and $(1-\lambda)B \subseteq C$ and both sets $\lambda A, (1-\lambda)B$ are disjoint (apart from $\{0\}$). Hence

$$\mathrm{Vol}_1(C) \geq \mathrm{Vol}_1(\lambda A) + \mathrm{Vol}_1((1-\lambda)B) \geq \mathrm{Vol}_1(A)^{\lambda}\cdot \mathrm{Vol}_1(B)^{1-\lambda}$$

using the inequality for Arithmetic Mean vs Geometric Mean. $\square$

Claim II. *BM III for $n = 1 \Rightarrow$ Prékopa-Leindler inequality for $n = 1$.*

Proof of claim. Take function $f, g, h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and fix some λ with $h(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} g(y)^{1-\lambda}$ for $x, y \in \mathbb{R}$. We may assume that the functions are normalized so that $f(x), g(x), h(x) \leq 1$. Observe that the assumption implies that if $f(x) \geq t$ and $g(y) \geq t$, then $h(\lambda x + (1-\lambda)y) \geq t$. In particular the *level sets* of the functions satisfy

$$\{x \mid h(x) \geq t\} \supseteq \lambda\cdot\{x \mid f(x) \geq t\} + (1-\lambda)\cdot\{x \mid g(x)\geq t\} \qquad (*)$$

for all t. Then

$$\begin{aligned}
\int_{\mathbb{R}} h(x)dx \quad &= \int_0^1 \mathrm{Vol}_1(\{x \mid h(x)\geq t\})dt \\
&\overset{(*)}{\geq} \int_0^1 \mathrm{Vol}_1\Big(\lambda\cdot\{x \mid f(x)\geq t\} + (1-\lambda)\cdot\{x\mid g(x)\geq t\}\Big)dt \\
&\overset{\text{BM I}}{\geq} \int_0^1 \Big(\lambda\cdot\mathrm{Vol}_1(\{x\mid f(x)\geq t\}) + (1-\lambda)\cdot \mathrm{Vol}_1(\{x \mid g(x)\geq t\})\Big)\,dt \\
&= \lambda\cdot\Big(\int_{\mathbb{R}} f(x)dx\Big) + (1-\lambda)\cdot\Big(\int_{\mathbb{R}} g(x)dx\Big) \\
&\overset{AMGM}{\geq} \Big(\int_{\mathbb{R}} f(x)dx\Big)^{\lambda}\cdot\Big(\int_{\mathbb{R}} g(x)dx\Big)^{1-\lambda} \qquad \square
\end{aligned}$$

Claim III. *The Prékopa-Leindler inequality holds for every dimension $n \geq 2$.*
Proof of claim. We prove the claim by induction over n. Fix $0 < \lambda < 1$ and functions $f, g, h : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $h(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq f(\boldsymbol{x})^\lambda g(\boldsymbol{y})^{1-\lambda}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We will write $\boldsymbol{x} = (\bar{\boldsymbol{x}}, x_n)$ with $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. We define 1-dimensional functions $F, G, H : \mathbb{R} \to \mathbb{R}_{\geq 0}$ by letting $F(x_n) := \int_{\mathbb{R}^{n-1}} f(\bar{\boldsymbol{x}}, x_n) d\boldsymbol{x}$ — similarly we define G and H. Let us also write $f_{x_n} : \mathbb{R}^{n-1} \to \mathbb{R}_{\geq 0}$ as the function with fixed last coordinate, i.e. $f_{x_n}(\bar{\boldsymbol{x}}) := f(\bar{\boldsymbol{x}}, x_n)$. Then the assumption carries over to

$$h_{\lambda x_n + (1-\lambda) y_n}(\lambda \bar{\boldsymbol{x}} + (1-\lambda)\bar{\boldsymbol{y}}) \geq f_{x_n}(\bar{\boldsymbol{x}})^\lambda \cdot g_{y_n}(\bar{\boldsymbol{y}})^{1-\lambda} \quad \forall x_n, y_n \in \mathbb{R}\ \forall \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \in \mathbb{R}^{n-1}$$

Then for a *fixed* pair (x_n, y_n) we can apply the $(n-1)$-dimensional Prékopa-Leindler inequality to derive that

$$\begin{aligned} H(\lambda x_n + (1-\lambda) y_n) &= \int_{\mathbb{R}^{n-1}} h_{\lambda x_n + (1-\lambda) y_n}(\bar{\boldsymbol{x}}) d\bar{\boldsymbol{x}} \\ &\overset{(n-1)\text{-dim. PL}}{\geq} \left(\int_{\mathbb{R}^{n-1}} f_{x_n}(\bar{\boldsymbol{x}}) d\bar{\boldsymbol{x}} \right)^\lambda \left(\int_{\mathbb{R}^{n-1}} g_{y_n}(\bar{\boldsymbol{y}}) d\bar{\boldsymbol{y}} \right)^{1-\lambda} \\ &= F(x_n)^\lambda \cdot G(y_n)^{1-\lambda} \end{aligned}$$

That means we have satisfied the assumptions to apply the 1-dimensional Prékopa-Leindler Inequality to the functions F, G, H and

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathbb{R}} H(x_n) dx_n \overset{\text{1-dim PL}}{\geq} \left(\int_{\mathbb{R}} F(x_n) dx_n \right)^\lambda \left(\int_{\mathbb{R}} G(x_n) dx_n \right)^{1-\lambda}$$

Claim IV. *Prékopa-Leindler inequality for dim. $n \Rightarrow$ BM III for dim. n.*
Proof of claim. Let $A, B \subseteq \mathbb{R}^n$ be measurable sets and let $0 < \lambda < 1$. We will use the characteristic functions $f := \mathbf{1}_A$, $g := \mathbf{1}_B$ and $h := \mathbf{1}_{\lambda A + (1-\lambda) B}$ of the involved sets. Now take $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We need to argue that

$$h(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \overset{!}{\geq} f(\boldsymbol{x})^\lambda g(\boldsymbol{y})^{1-\lambda} = \begin{cases} 1 & \text{if } \boldsymbol{x} \in A \text{ and } \boldsymbol{y} \in B \\ 0 & \text{otherwise} \end{cases}$$

There is only something to show if the right hand side is 1 — then $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in (\lambda A + (1-\lambda) B)$ and the left hand side is 1 as well. Either way, we can apply the Prékopa-Leindler inequality to get

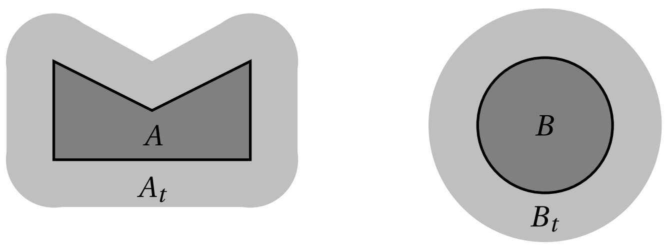
$$\begin{aligned} \mathrm{Vol}_n(\lambda A + (1-\lambda) B) &= \int_{\mathbb{R}^n} h(\boldsymbol{z}) d\boldsymbol{z} \overset{\text{PL}}{\geq} \left(\int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x} \right)^\lambda \left(\int_{\mathbb{R}^n} g(\boldsymbol{y}) d\boldsymbol{y} \right)^{1-\lambda} \\ &= \mathrm{Vol}_n(A)^\lambda \cdot \mathrm{Vol}_n(B)^{1-\lambda} \end{aligned}$$ □

Putting everything together, this proves both, the Brunn-Minkowski Inequality and the Prékopa-Leindler Inequality.

1.8.2 The isoperimetric inequality

One of the few facts in convex geometry that are more widely known to non-mathematicians is that among bodies with identical volumes, the Euclidean ball minimizes the surface area. We abbreviate $d(\boldsymbol{x}, A) := \inf\{\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{y} \in A\}$ as the distance of a point $\boldsymbol{x}$ to a set A and we denote $A_t := \{\boldsymbol{x} \in \mathbb{R}^n \mid d(\boldsymbol{x}, A) \leq t\}$ as its *t-neighborhood*. If A is compact, then $A_t = A + tB_2^n$.

Theorem 1.44 (Isoperimetric Inequality)**.** *Let $A \subseteq \mathbb{R}^n$ be a compact set and let $B := rB_2^n$ be the Euclidean ball so that $Vol_n(A) = Vol_n(B)$. Then $Vol_n(A_t) \geq Vol_n(B_t)$ for any $t \geq 0$.*



Proof. After rescaling both sets, suppose that $B = B_2^n$ is the unit radius ball. Using the Brunn-Minkowski inequality we can bound

$$\begin{aligned} \mathrm{Vol}_n(A_t) &= \mathrm{Vol}_n(A+tB) \overset{\text{BM I}}{\geq} \left(\mathrm{Vol}_n(A)^{1/n} + t\cdot \mathrm{Vol}_n(B)^{1/n}\right)^n \\ &\overset{\mathrm{Vol}_n(A)=\mathrm{Vol}_n(B)}{=} (1+t)^n \cdot \mathrm{Vol}_n(B) = \mathrm{Vol}_n(B_t) \end{aligned}$$

□

Then notice that the surface area of a compact set can be defined as

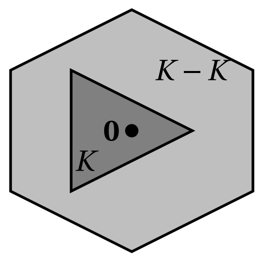
$$\mathrm{Vol}_{n-1}(\partial A) := \lim_{\varepsilon \to 0} \frac{\mathrm{Vol}_n(A+\varepsilon B_2^n) - \mathrm{Vol}_n(A)}{\varepsilon}$$

hence B is the Euclidean ball with $\mathrm{Vol}_n(A) = \mathrm{Vol}_n(B)$, then indeed $\mathrm{Vol}_{n-1}(\partial A) \geq \mathrm{Vol}_{n-1}(\partial B)$.

1.9 Polar coordinates and the inequality of Rogers and Shephard

In many settings, it is desirable to have a convex body K that is also symmetric. For example if we have symmetry, then we know that $\|\cdot\|_K$ is a norm. Also, we will later see that covering numbers are easier to handle for symmetric bodies. That leads to the natural question whether there is a generic procedure of approximating an asymmetric convex body K with a symmetric body. In the next Chapter on John's Theorem we will see that there is always an ellipsoid $\mathcal{E}$ that after proper translation of K satisfies $\frac{1}{n}\mathcal{E} \subseteq K \subseteq \mathcal{E}$. In terms of the volume this guarantees a bound of $\mathrm{Vol}_n(\mathcal{E}) \leq n^n \cdot \mathrm{Vol}_n(K)$. It turns out that there is a better approximation of K with a symmetric body if we drop the first inclusion requirement.

For a convex body $K \subseteq \mathbb{R}^n$ we define $K - K := \{\boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{y} \in K\}$ as the *difference body*. Note that by construction $K - K$ is centrally symmetric even if K was not. Moreover, if $\mathbf{0} \in K$, then $K \subseteq K - K$.

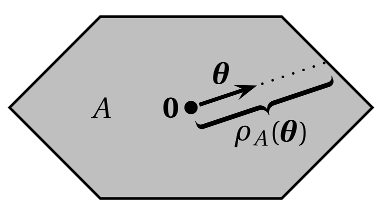


We will next prove the inequality of Rogers and Shephard showing that $K - K$ is not much bigger than K in the sense that $\mathrm{Vol}_n(K - K) \leq 2^{2n} \cdot \mathrm{Vol}_n(K)$. We should also remark that by Brunn-Minkowski we also know a lower bound of $\mathrm{Vol}_n(K - K) = 2^n \cdot \mathrm{Vol}_n(\frac{1}{2}K + \frac{1}{2}(-K)) \geq 2^n \cdot \mathrm{Vol}_n(K)$ for any convex body K.

Recall that the *radial function* of a convex set A is defined by

$$\rho_A(\boldsymbol{\theta}) := \max\{t \geq 0 \mid t \cdot \boldsymbol{\theta} \in A\} \stackrel{\text{if } A \text{ symmetric}}{=} \frac{1}{\|\boldsymbol{\theta}\|_A} \quad \forall \boldsymbol{\theta} \in S^{n-1}$$

To see the connection between radial function and Minkowski norm, note that if $\rho_A(\boldsymbol{\theta}) = t$, then $\|t \cdot \boldsymbol{\theta}\|_A = 1$ and so $\|\boldsymbol{\theta}\|_A = \frac{1}{t}$. Geometrically speaking, $\rho_A(\boldsymbol{\theta})$ is the distance one has to walk from the origin in direction $\boldsymbol{\theta}$ until exiting the body.



We can use the radial function to rewrite an integral to an integral in *polar coordinates.*

Theorem 1.45 (Integration in polar coordinates)**.** *For any integrable function $f : \mathbb{R}^n \to \mathbb{R}$ and a convex body $A \subseteq \mathbb{R}^n$ with $\mathbf{0} \in A$ we have*

$$\int_A f(\boldsymbol{x})\, d\boldsymbol{x} = Vol_{n-1}(S^{n-1}) \cdot \mathop{\mathbb{E}}_{\boldsymbol{\theta} \in S^{n-1}} \Big[\int_0^{\rho_A(\boldsymbol{\theta})} f(r \cdot \boldsymbol{\theta}) \cdot r^{n-1}\, dr \Big]$$

This also implies a convinient formula to express the volume of a body:

Lemma 1.46 (Volume in polar coordinates)**.** *For a convex body $A \subseteq \mathbb{R}^n$ with $\mathbf{0} \in A$ one has $Vol_n(A) = Vol_n(B_2^n) \cdot \mathbb{E}_{\boldsymbol{\theta} \in S^{n-1}}[\rho_A(\boldsymbol{\theta})^n]$.*

Proof. Integrating the characteristic function of A in polar coordinates gives

$$\mathrm{Vol}_n(A) = \int_A 1 d\boldsymbol{x} = \underbrace{\mathrm{Vol}_{n-1}(S^{n-1})}_{=n\mathrm{Vol}_n(B_2^n)} \cdot \mathop{\mathbb{E}}_{\boldsymbol{\theta} \in S^{n-1}} \Big[\underbrace{\int_0^{\rho_A(\boldsymbol{\theta})} r^{n-1} dr}_{=\frac{1}{n}\rho_A(\boldsymbol{\theta})^n} \Big] = \mathrm{Vol}_n(B_2^n) \mathop{\mathbb{E}}_{\boldsymbol{\theta} \in S^{n-1}} \big[\rho_A(\boldsymbol{\theta})^n\big]$$

□

Theorem 1.47 (Rogers-Shephard)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then*

$$2^n \cdot Vol_n(K) \leq Vol_n(K - K) \leq 2^{2n} \cdot Vol_n(K).$$

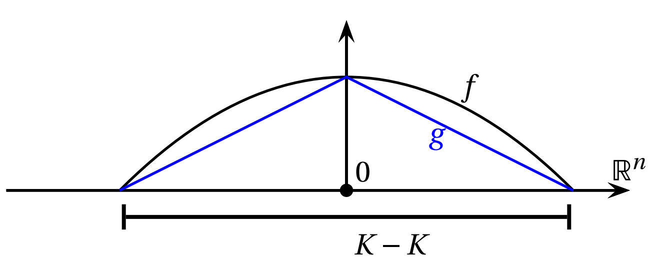
Proof. We already argued the lower bound. To keep the calculations short we will prove a slightly weaker upper bound of $4^{2n}/n$. More precisely we will prove the two (in)equalities (I) and (II) in

$$\frac{n}{4^{2n}} \cdot \mathrm{Vol}_n(K) \overset{(I)}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x} \in K-K} \big[\mathrm{Vol}_n(K \cap (\boldsymbol{x} + K))\big] \overset{(II)}{=} \frac{\mathrm{Vol}_n(K)^2}{\mathrm{Vol}_n(K-K)} \leq \frac{1}{2^n} \mathrm{Vol}_n(K) \quad (*)$$

Then rearranging gives a bound of $\mathrm{Vol}_n(K - K) \leq \frac{4^{2n}}{n} \cdot \mathrm{Vol}_n(K)$. We define the function $f : K - K \to \mathbb{R}_{\geq 0}$ with

$$\begin{aligned} f(\boldsymbol{x}) \quad &:= \quad \mathrm{Vol}_n(K \cap (\boldsymbol{x} + K))^{1/n} = \mathrm{Vol}_n\big(\{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y} \in K \text{ and } \boldsymbol{y} - \boldsymbol{x} \in K\}\big)^{1/n} \quad (**) \\ &= \quad \mathrm{Vol}_n\Big(\Big\{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y} - \frac{\boldsymbol{x}}{2} \in K \text{ and } \boldsymbol{y} + \frac{\boldsymbol{x}}{2} \in K\Big\}\Big)^{1/n} \end{aligned}$$

From the last representation in $(**)$ we can also see that the function is symmetric with $f(\boldsymbol{x}) = f(-\boldsymbol{x})$. Observe that if $f(\boldsymbol{x}) > 0$ then there is a $\boldsymbol{y}$ with $\boldsymbol{y} \in K$ and $\boldsymbol{y} - \boldsymbol{x} \in K$ and hence $\boldsymbol{x} = \boldsymbol{y} - (\boldsymbol{y} - \boldsymbol{x}) \in K - K$. So, indeed the support of f is contained in $K - K$ with $f(\boldsymbol{x}) = 0$ for points $\boldsymbol{x}$ on the boundary of $K - K$. We know from Lemma 1.34 that the function f is concave. Since $f(\mathbf{0}) = \mathrm{Vol}_n(K)^{1/n}$, we can use these insights to define a lower bound function $g : K - K \to \mathbb{R}_{\geq 0}$ with $f(\boldsymbol{x}) \geq g(\boldsymbol{x})$ for all $\boldsymbol{x} \in K - K$ by letting $g(r \cdot \boldsymbol{\theta}) := \mathrm{Vol}_n(K)^{1/n} \cdot (1 - \frac{r}{\rho_{K-K}(\boldsymbol{\theta})})$.



Note that geometrically, for a fixed direction $\boldsymbol{\theta}$, the function g is linear in r. Then using integration in polar coordinates we obtain

$$\begin{aligned}
& \int_{K-K} \mathrm{Vol}_n(K \cap (\boldsymbol{x}+K))\, d\boldsymbol{x} \\
\overset{f(\boldsymbol{x})\geq g(\boldsymbol{x})}{\geq} & \int_{K-K} g(\boldsymbol{x})^n \, d\boldsymbol{x} \\
\overset{\text{Thm. 1.45}}{=} & n\mathrm{Vol}_n(B_2^n) \cdot \mathop{\mathbb{E}}_{\boldsymbol{\theta}\in S^{n-1}} \Big[\int_0^{\rho_{K-K}(\boldsymbol{\theta})} g(r\boldsymbol{\theta})^n \cdot r^{n-1}\, dr \Big] \\
\overset{\text{Def } g}{=} & n\mathrm{Vol}_n(B_2^n) \cdot \mathrm{Vol}_n(K) \mathop{\mathbb{E}}_{\boldsymbol{\theta}\in S^{n-1}} \Big[\int_0^{\rho_{K-K}(\boldsymbol{\theta})} \Big(1 - \frac{r}{\rho_{K-K}(\theta)}\Big)^n \cdot r^{n-1} dr \Big] \\
\geq \; & n\mathrm{Vol}_n(B_2^n) \cdot \mathrm{Vol}_n(K) \mathop{\mathbb{E}}_{\boldsymbol{\theta}\in S^{n-1}} \Big[\int_{\frac{1}{4}\rho_{K-K}(\boldsymbol{\theta})}^{\frac{3}{4}\rho_{K-K}(\boldsymbol{\theta})} \underbrace{\Big(1 - \frac{r}{\rho_{K-K}(\theta)}\Big)^n}_{\geq (1/4)^n} \cdot \underbrace{r^{n-1}}_{\geq (\frac{1}{4}\rho_{K-K}(\boldsymbol{\theta}))^{n-1}} dr \Big] \\
\geq \; & n \cdot (1/4)^{2n} \cdot \mathrm{Vol}_n(K) \cdot \underbrace{\mathrm{Vol}_n(B_2^n) \mathop{\mathbb{E}}_{\boldsymbol{\theta}\in S^{n-1}} \big[\rho_{K-K}(\boldsymbol{\theta})^n\big]}_{=\mathrm{Vol}_n(K-K) \text{ by Lem 1.46}}
\end{aligned}$$

This shows (I). To show (II) we can use Fubini's Theorem to swap the integration order and get

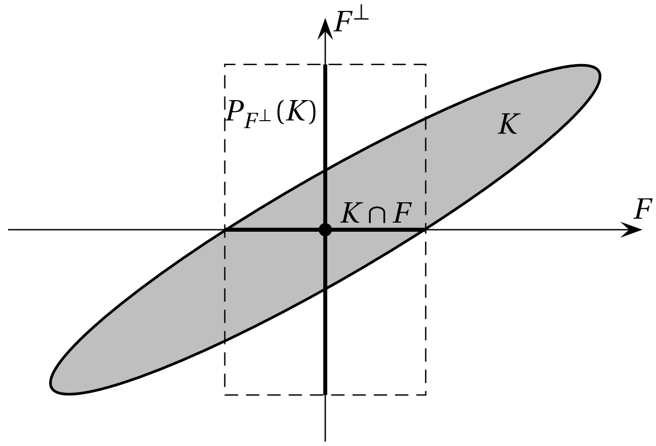
$$\int_{K-K} \mathrm{Vol}_n(K \cap (\boldsymbol{x}+K))\, d\boldsymbol{x} = \mathrm{Vol}_{2n}\big(\{(\boldsymbol{x},\boldsymbol{y}) \mid \boldsymbol{y} \in K \text{ and } \boldsymbol{y}-\boldsymbol{x} \in K\}\big) = \mathrm{Vol}_n(K)^2$$

That settles (II) and hence the Theorem. $\square$

We will also show an inequality that has a proof-strategy similar to Rogers-Shephard, even if the statements sound quite different. Recall that for a subspace $F \subseteq \mathbb{R}^n$, $F^\perp$ is the complementary subspace and $P_{F^\perp} : \mathbb{R}^n \to F^\perp$ is the *orthogonal projection* into the orthogonal complement of F. Note that for an arbitrary k-dimensional subspace F, neither $\mathrm{Vol}_k(K \cap F)$ nor $\mathrm{Vol}_{n-k}(P_{F^\perp}(K))$ alone gives much information on the size of K. But surprisingly it turns out that the *product* of $\mathrm{Vol}_k(K \cap F)$ and $\mathrm{Vol}_{n-k}(P_{F^\perp}(K))$ is a very good proxy for $\mathrm{Vol}_n(K)$. Note that the factor of 2^{-n} in the following estimate can be improved to $\binom{n}{k}^{-1}$ by doing the calculations more carefully. As usually we prefer to keep the exposition simple.

Lemma 1.48. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $F \subseteq \mathbb{R}^n$ be a subspace with $k := \dim(F)$. Then*

$$2^{-n} \leq \frac{Vol_n(K)}{Vol_k(K \cap F) \cdot Vol_{n-k}(P_{F^\perp}(K))} \leq 1$$



Proof. We define the function $G(\boldsymbol{x}) := \mathrm{Vol}_k(K \cap (F+\boldsymbol{x}))^{1/k}$. Then by *Brunn's Concavity Principle* (Theorem 1.32), we know that $G(\boldsymbol{x})$ is concave. As K is symmetric we know that also G is symmetric, meaning that $G(-\boldsymbol{x}) = G(\boldsymbol{x})$ for all $\boldsymbol{x}$. From this it follows that G is maximized for $\boldsymbol{x} = \mathbf{0}$. Then

$$\mathrm{Vol}_n(K) \stackrel{\text{Fubini}}{=} \int_{\boldsymbol{x} \in P_{F^\perp}(K)} \underbrace{G(\boldsymbol{x})^k}_{\leq G(\mathbf{0})^k} d\boldsymbol{x} \leq \mathrm{Vol}_{n-k}(P_{F^\perp}(K)) \cdot \mathrm{Vol}_k(K \cap F)$$

which shows the upper bound. For the lower bound, the crucial observation is that by the concavity of G we know that for every $\boldsymbol{x} \in \frac{1}{2}P_{F^\perp}(K)$ one has $G(\boldsymbol{x}) \geq \frac{1}{2}G(\mathbf{0})$. Then only counting that part of the volume gives

$$\begin{aligned}
\mathrm{Vol}_n(K) &\geq \int_{\boldsymbol{x} \in \frac{1}{2}P_{F^\perp}(K)} G(\boldsymbol{x})^k d\boldsymbol{x} \geq \int_{\boldsymbol{x} \in \frac{1}{2}P_{F^\perp}(K)} \left(\frac{1}{2}G(\mathbf{0})\right)^k d\boldsymbol{x} \\
&= (1/2)^{n-k} \cdot \mathrm{Vol}_{n-k}(P_{F^\perp}(K)) \cdot (1/2)^k \cdot \mathrm{Vol}_k(K \cap F)
\end{aligned}$$

This finishes the claim. $\square$

1.10 Exercises

Exercise 1.1.

Prove that for any symmetric convex body $K \subseteq \mathbb{R}^n$ and any k-dimensional subspace F one has $\mathrm{Vol}_n(K) \geq \left(1 - \frac{k}{n}\right)^{n-k} \left(\frac{k}{n}\right)^k \cdot \mathrm{Vol}_k(K \cap F) \cdot \mathrm{Vol}_{n-k}(P_{F^\perp}(K))$.

Exercise 1.2.

Let $\mathcal{K}$ be the set of symmetric convex bodies in $\mathbb{R}^n$ and abbreviate $\Delta(K,Q) := \ln(d_{BM}(K,Q))$. Prove that Δ is a pseudometric on $\mathcal{K}$.

Exercise 1.3.
Let $\mathcal{S}_{n\times n} := \{\boldsymbol{A} \in \mathbb{R}^{n\times n} \mid \boldsymbol{A}^T = \boldsymbol{A}\}$ be the vector space of symmetric matrices. Let $\lambda_i(\boldsymbol{A})$ be the ith Eigenvalue of $\boldsymbol{A}$. Consider $\mathcal{B}_\infty^{n\times n} = \{\boldsymbol{A} \in \mathcal{S}_{n\times n} \mid |\lambda_i(\boldsymbol{A})| \leq 1 \;\; \forall i \in [n]\}$ and $\mathcal{B}_1^{n\times n} := \{\boldsymbol{A} \in \mathcal{S}_{n\times n} \mid \sum_{i=1}^n |\lambda_i(\boldsymbol{A})| \leq 1\}$. Prove that $(\mathcal{B}_\infty^{n\times n})^\circ = \mathcal{B}_1^{n\times n}$ (using $\langle \boldsymbol{A},\boldsymbol{B}\rangle_F := \sum_{i=1}^n \sum_{j=1}^n A_{ij}B_{ij}$ as inner product).

Exercise 1.4.
Recall that the Generalized Cauchy Schwarz inequality says that $|\langle \boldsymbol{x},\boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|_K \cdot \|\boldsymbol{y}\|_{K^\circ}$ for a symmetric convex body $K \subseteq \mathbb{R}^n$ and $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$.

a) Show that for any parameter $\rho \geq 1$ there is a symmetric convex body $K \subseteq \mathbb{R}^2$ and $\boldsymbol{x} \in \mathbb{R}^2$ so that $\rho \langle \boldsymbol{x},\boldsymbol{x}\rangle \leq \|\boldsymbol{x}\|_K \cdot \|\boldsymbol{x}\|_{K^\circ}$.
b) Show that for any symmetric convex body $K \subseteq \mathbb{R}^n$ there exists at least one non-zero $\boldsymbol{x}^* \in K$ so that $\|\boldsymbol{x}^*\|_2^2 = \|\boldsymbol{x}^*\|_K \cdot \|\boldsymbol{x}^*\|_{K^\circ}$.

Exercise 1.5.
Let $p,q \in [1,\infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Prove that $(B_p^n)^\circ = B_q^n$.

Exercise 1.6.
Prove the following Theorem of Bieberbach from 1915: Let $K \subseteq \mathbb{R}^n$ be a compact set. Then

$$\mathrm{Vol}_n(K) \leq \mathrm{Vol}_n(B_2^n) \cdot \frac{\mathrm{diam}(K)^n}{2^n}$$

where $\mathrm{diam}(K) := \max\{\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{x},\boldsymbol{y} \in K\}$.
Hint. Use Steiner symmetrization.

Exercise 1.7.
In this exercise we want to give a proof for Urysohn's inequality using a different symmetrization strategy. You may use the following fact without proof: *For every convex body K and $\varepsilon > 0$ there is an $N := N(K,\varepsilon)$ and orthogonal transformations $U_1,\dots,U_N : \mathbb{R}^n \to \mathbb{R}^n$ so that the body $Q := \frac{1}{N}(U_1(K)+\ldots+U_N(K))$ satisfies $(1-\varepsilon)RB_2^n \subseteq Q \subseteq (1+\varepsilon)RB_2^n$ for some R.* Show the following for any convex body $K \subseteq \mathbb{R}^n$ (without the help of Steiner's symmetrization):

(i) Any body Q as above has $w(Q) = w(K)$.
(ii) Any body Q as above has $\mathrm{Vol}_n(Q) \geq \mathrm{Vol}_n(K)$.
(iii) Any convex body K has $w(K) \geq 2(\frac{\mathrm{Vol}_n(K)}{\mathrm{Vol}_n(B_2^n)})^{1/n}$.

Comment. One can prove the fact mentioned above by picking independent random orthogonal transformations $U_1, \dots, U_N$ for N large enough. But this is not part of the exercise.

Exercise 1.8.
Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and consider a strip $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid |\langle \boldsymbol{x}, \boldsymbol{\theta} \rangle| \leq s\}$ with $\boldsymbol{\theta} \in S^{n-1}$. Prove that $\mathrm{Vol}_n(K \cap S) \geq \mathrm{Vol}_n(K) \cdot \frac{1}{n}(1 - (1 - \frac{s}{h_K(\boldsymbol{\theta})})^n)$.

Chapter 2

John's Theorem

In this chapter, we will present a tremendously useful result of John with numerous applications.

> *For any convex body $K \subseteq \mathbb{R}^n$, there is an ellipsoid $\mathcal{E}$ so that after proper translation $\mathcal{E} \subseteq K \subseteq n\mathcal{E}$.*

In terms of the Banach Mazur distance, this means that $d_{BM}(K, B_2^n) \leq n$ for any convex body K. In many settings, John's Theorem can give a quick estimate for the desired quantity. For example, for the inequality of Rogers-Shephard, after translation we can find an ellipsoid with $\mathcal{E} \subseteq K \subseteq n\mathcal{E}$ and hence $K - K \subseteq 2n\mathcal{E}$. Then we can conclude that $\mathrm{Vol}_n(K-K) \leq \mathrm{Vol}_n(2n\mathcal{E}) \leq (2n)^n \cdot \mathrm{Vol}_n(K)$. We know that this is not a tight bound, but helpful to understand the ball park of what might be possible.

John's Theorem can be sharpened for symmetric sets:

> *For any symmetric convex body $K \subseteq \mathbb{R}^n$, there is an ellipsoid $\mathcal{E}$ so that $\mathcal{E} \subseteq K \subseteq \sqrt{n} \cdot \mathcal{E}$.*

One consequence of this bound in particular is that for *any* norm $\|\cdot\|_K$ in $\mathbb{R}^n$, there is a matrix $\boldsymbol{A}$ so that $\|\boldsymbol{x}\|_K \leq \|\boldsymbol{A}\boldsymbol{x}\|_2 \leq \sqrt{n} \cdot \|\boldsymbol{x}\|_K$ for all $\boldsymbol{x} \in \mathbb{R}^n$. The Theorem of John is excellently described in the wonderful survey of Ball [Bal97] and we refer to it for more details.

We already mentioned ellipsoids without providing a formal definition so far. A linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ can be uniquely identified with the underlying matrix. By a slight abuse of notation we write $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ as the matrix so that $A(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. Then for a bijective linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ we call the set $A(B_2^n) = \{\boldsymbol{A}\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n \text{ with } \|\boldsymbol{x}\|_2 \leq 1\}$ an *ellipsoid*. Note that in our notation, an ellipsoid is always an origin-centered convex body.

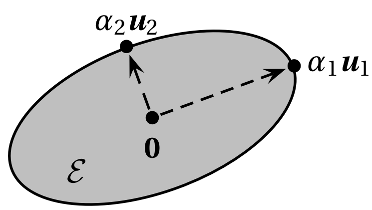
We want to comment on the matrix representing the ellipsoid. Consider the *singular value decomposition* $\boldsymbol{A} = \sum_{i=1}^n \alpha_i \boldsymbol{u}_i \boldsymbol{v}_i^T$ where $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ and $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ are two orthonormal bases and $\alpha_1, \dots, \alpha_n > 0$. Then

$$A(B_2^n) = \left\{ \sum_{i=1}^n \alpha_i \boldsymbol{u}_i \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle \mid \|\boldsymbol{x}\|_2 \leq 1 \right\} \overset{\text{symmetry of } B_2^n}{=} \left\{ \sum_{i=1}^n \alpha_i \boldsymbol{u}_i \langle \boldsymbol{u}_i, \boldsymbol{y} \rangle \mid \|\boldsymbol{y}\|_2 \leq 1 \right\} = B(B_2^n)$$

where $\boldsymbol{B} := \sum_{i=1}^n \alpha_i \boldsymbol{u}_i \boldsymbol{u}_i^T$ is a symmetric PSD matrix. In other words, the matrix $\boldsymbol{A}$ with $\mathcal{E} = \{\boldsymbol{A}\boldsymbol{x} \mid \boldsymbol{x} \in B_2^n\}$ is not unique but it can always be chosen to be a symmetric positive definite matrix. Now suppose that $\boldsymbol{B}$ is indeed the PSD matrix as defined above. Clearly $\mathrm{Vol}_n(\mathcal{E}) = |\det(\boldsymbol{B})| \cdot \mathrm{Vol}_n(B_2^n)$. Moreover,

$$\mathcal{E} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \sum_{i=1}^n \frac{\langle \boldsymbol{x}, \boldsymbol{u}_i \rangle^2}{\alpha_i^2} \leq 1 \right\} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^T \boldsymbol{B}^{-2} \boldsymbol{x} \leq 1 \right\} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{B}^{-1}\boldsymbol{x}\|_2^2 \leq 1 \right\}$$

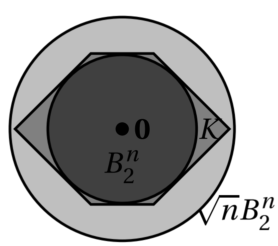
is an alternative representation that has a clean geometric interpretation: the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ are the *axes* of the ellipsoid and α_i is the length of the ith axis.



2.1 The most basic John's Theorem

We will see that indeed the largest volume ellipsoid inside K has the property of John's Theorem. For proofs it will be easiest to argue if we apply a linear transformation so that that ellipsoid is the unit ball B_2^n. For that purpose, we say that a convex body K is in *John position* if B_2^n is an ellipsoid of maximum volume contained in K (even if translations are allowed). The concept of a *position* is also used in other contexts in convex geometry. The idea is always to apply a linear transformation to a convex body so that some constraint is satisfied.

Theorem 2.1 (John's Theorem for Symmetric Bodies [Joh48])**.** *Let K be a centrally symmetric convex body in John position. Then $B_2^n \subseteq K \subseteq \sqrt{n} B_2^n$.*



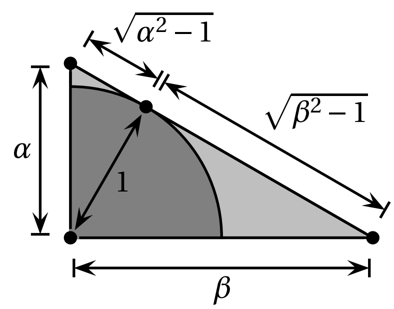
Proof. We have $B_2^n \subseteq K$ by assumption. Now suppose by symmetry reasons that $R\boldsymbol{e}_1 \in K$ where $R > \sqrt{n}$. Then by symmetry $\mathrm{conv}\{B_2^n \cup \{\pm R\boldsymbol{e}_1\}\} \subseteq K$. We consider the ellipsoid

$$\mathcal{E} := \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \frac{x_1^2}{a^2} + \sum_{i=2}^n \frac{x_i^2}{b^2} \leq 1\right\}$$

obtained by stretching the ball by a factor $a > 1$ in direction $\boldsymbol{e}_1$ and shrinking it slightly to a factor $b < 1$ in all other directions. We pick b maximal so that the ellipsoid just touches the body $\mathrm{conv}\{B_2^n \cup \{\pm R\boldsymbol{e}_1\}\} \subseteq K$ from the inside. It remains to determine what constellations of a and b are feasible. Note that by symmetry it suffices to consider the situation in a 2-dimensional plane spanned by $\boldsymbol{e}_1$ and one orthogonal vector. First we need a simple fact:

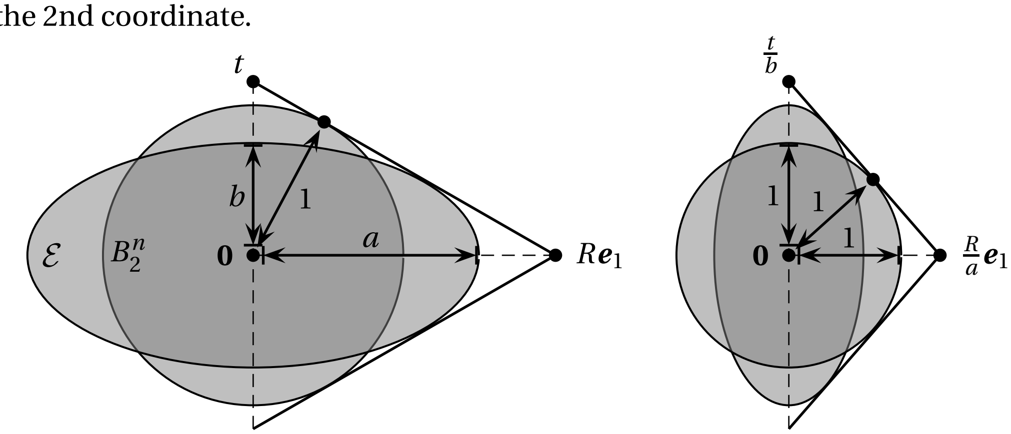
Claim. *Suppose that the triangle with* $(\alpha, 0), (0, \beta), (0, 0) \in \mathbb{R}^2_{\geq 0}$ *is touching* B_2^2. *Then* $\beta = \frac{\alpha}{\sqrt{\alpha^2-1}}$.

Proof of claim. Consider the following figure:



We derive that the hypothenuse has length $\sqrt{\alpha^2+\beta^2} = \sqrt{\alpha^2-1} + \sqrt{\beta^2-1}$ which can be rearranged to $\beta = \frac{\alpha}{\sqrt{\alpha^2-1}}$. □

Now back to John's Theorem. Consider the following two figures, where the 2nd one is obtained by shrinking the first one by a in the 1st coordinate and b in the 2nd coordinate.



Using the claim we can infer from the 1st picture that $t^2 = \frac{R^2}{R^2-1}$ and from the 2nd picture we infer that $(\frac{t}{b})^2 = \frac{(R/a)^2}{(R/a)^2-1}$. Substituting t and rearranging gives

$$a^2 = b^2 + R^2(1-b^2) \quad (*)$$

Now suppose that $R = \sqrt{n+\delta}$ for a fixed $\delta > 0$. If we set $b := 1-\varepsilon$, then we can rearrange $(*)$ to $a = \sqrt{1+\varepsilon\cdot 2\cdot(R^2-1)+\varepsilon^2\cdot(1-R^2)} = 1+(n-1+\delta)\cdot\varepsilon - \Theta(\varepsilon^2)$ where the O-notation hides factors dependent on n and δ. Moreover, the volume changes as

$$\frac{\mathrm{Vol}_n(\mathcal{E})}{\mathrm{Vol}_n(B_2^n)} = ab^{n-1} = (1+(n-1+\delta)\varepsilon - \Theta(\varepsilon^2))\cdot(1-\varepsilon)^{n-1} = 1+\delta\varepsilon \pm O(\varepsilon^2) \overset{\varepsilon \text{ small}}{>} 1$$

$\square$

One can check that the bound of $\sqrt{n}$ is tight, see for example the cube $K = [-1,1]^n$ that indeed contains B_2^n as largest volume ellipsoid. We will also state the version of John's Theorem for non-symmetric bodies. Again the bound will be tight, see for example the simplex.

Theorem 2.2 (John's Theorem for Asymmetric Bodies)**.** *Let $K \subseteq \mathbb{R}^n$ be any convex body in John position. Then $B_2^n \subseteq K \subseteq nB_2^n$.*

The proof works similar to the symmetric case — if there is a point $R\boldsymbol{e}_1 \in K$ with $R > n$, then stretch the ball B_2^n into direction $\boldsymbol{e}_1$ and shrink it in all orthogonal directions. The only difference is that now we also need to move the center towards $\boldsymbol{e}_1$. We will skip the formalities here.

The maximum volume ellipsoid is always unique. We sketch the argument but refrain from a formal proof:

Lemma 2.3. *Every convex body $K \subseteq \mathbb{R}^n$ has a* unique *maximum volume ellipsoid $\mathcal{E} \subseteq K$.*

Proof. Existence follows from compactness arguments. For uniqueness, suppose that there are two ellipsoids $\mathcal{E}_A = \boldsymbol{a} + A(B_2^n)$ and $\mathcal{E}_B = \boldsymbol{b} + B(B_2^n)$ contained in K that have identical volume, say $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$ as we can pick the matrices to be PSD. If it happens that $\boldsymbol{A} = \boldsymbol{B}$ then we can define an ellipsoid with center $\frac{\boldsymbol{a}+\boldsymbol{b}}{2}$ that is slightly stretched in direction $\boldsymbol{a}-\boldsymbol{b}$ and hence has a larger volume. So suppose that $\boldsymbol{A} \neq \boldsymbol{B}$. Then one can prove the following general inequality:
Claim. *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ be positive definite matrices. Then* $\det(\boldsymbol{A}+\boldsymbol{B})^{1/n} \geq \det(\boldsymbol{A})^{1/n} + \det(\boldsymbol{B})^{1/n}$. *Equality holds iff the matrices are scalars of each other.*
We skip the proof which consists of a smart way to integrate plus an application

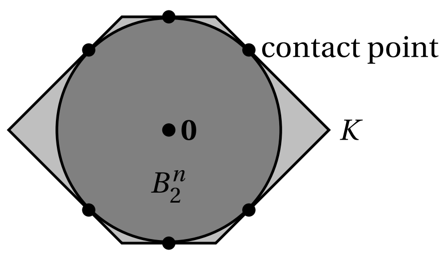
of Hölder's inequality. We can consider the ellipsoid obtained by averaging the matrices[1]:

$$\mathcal{E}' := \frac{\boldsymbol{a}+\boldsymbol{b}}{2} + \left\{\frac{1}{2}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x} \mid \boldsymbol{x} \in B_2^n\right\}$$

From the definition and convexity we see that $\mathcal{E}' \subseteq \frac{1}{2}(\mathcal{E}_A + \mathcal{E}_B) \subseteq K$. From the claim we see that $\mathrm{Vol}_n(\frac{1}{2}(\mathcal{E}_A + \mathcal{E}_B)) > \mathrm{Vol}_n(\mathcal{E}_A) = \mathrm{Vol}_n(\mathcal{E}_B)$ which then is a contradiction. □

2.2 Contact points

If we consider a convex body K in John's position, then it is not hard to see that some points on the boundary of K must be touching the boundary of the ball B_2^n — otherwise would could have scaled B_2^n to obtain a larger ellipsoid. Formally we call a point $\boldsymbol{x}$ a *contact point* of K is $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_K = 1$.



The next observation is that there must be contact points in "all directions" since otherwise we would have freedom to scale B_2^n orthogonal to the contact points and again obtain a larger ellipsoid. In fact, we can formalize nicely what it means that there are contact points in "all directions":

Theorem 2.4 (John)**.** *Let B_2^n be the maximum volume ellipsoid in a symmetric convex body $K \subseteq \mathbb{R}^n$. Then there are contact points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ of K and B_2^n for $m \leq \binom{n+1}{2} + 1 \leq n^2 + 1$ and scalars $c_1, \dots, c_m > 0$ so that*

$$\boldsymbol{I}_n = \sum_{j=1}^{m} c_j \boldsymbol{x}_j \boldsymbol{x}_j^T \quad (*)$$

In particular $\sum_{j=1}^m c_j = n$.

Proof. The condition $\sum_{j=1}^m c_j = n$ follows by just taking the trace of both sides of $(*)$, so we do not need to further discuss it. Let us abbreviate U as all the contact points. Consider

$$\mathcal{C} := \{\boldsymbol{u}\boldsymbol{u}^T \mid \boldsymbol{u} \in U\}$$

[1] It may be pointed out that in general the Minkowski sum of two ellipsoids is NOT an ellipsoid.

which is the set of outer products of contact points. Suppose for the sake of contradiction that no proper coefficients c_j exist. Then $\frac{\boldsymbol{I}_n}{n} \notin \text{conv}(\mathcal{C})$ and by the *Hyperplane Separation Theorem* (Theorem 1.22), there is a linear function $\Phi : \mathbb{R}^{n\times n} \to \mathbb{R}$ so that

$$\Phi\left(\frac{\boldsymbol{I}_n}{n}\right) < r \leq \Phi(\boldsymbol{x}\boldsymbol{x}^T) \quad \forall \boldsymbol{x} \in U \quad (**)$$

for some $r \in \mathbb{R}$. Recall that the function will be of the form $\Phi(\boldsymbol{X}) = \langle \boldsymbol{B}, \boldsymbol{X}\rangle$ where $\langle \cdot, \cdot \rangle$ is the *Frobenius inner product*. In fact, the matrices on both sides of $(**)$ are symmetric and hence we may also choose $\boldsymbol{B}$ is symmetric. Moreover, the matrices on both sides of $(**)$ lie all on the hyperplane $\langle \boldsymbol{I}_n, \boldsymbol{X}\rangle = 1$ which implies that one can subtract a multiple of the "normal vector" $\boldsymbol{B}$ to assume that $\langle \boldsymbol{B}, \boldsymbol{I}_n\rangle = 0$. We summarize that the hyperplane satisfies $\langle \boldsymbol{B}, \boldsymbol{I}_n\rangle = 0 < \langle \boldsymbol{B}, \boldsymbol{u}\boldsymbol{u}^T\rangle$ for all $\boldsymbol{u} \in U$. Consider the ellipsoid

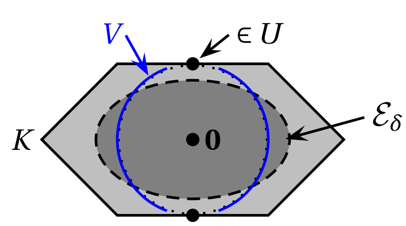
$$\mathcal{E}_\delta := \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle (\boldsymbol{I}_n + \delta \boldsymbol{B})\boldsymbol{x}, \boldsymbol{x}\rangle \leq 1\right\}$$

for some tiny enough $\delta > 0$. Note that the volume of the ellipsoid satisfies

$$\frac{\text{Vol}_n(\mathcal{E}_\delta)}{\text{Vol}_n(B_2^n)} = \frac{1}{\det(\boldsymbol{I}_n + \delta \boldsymbol{B})^{1/2}} \geq \frac{1}{(\frac{1}{n}\text{Tr}[\boldsymbol{I}_n + \delta \boldsymbol{B}])^{n/2}} = 1$$

where we use the inequality of the arithmetic-vs-geometric-mean to obtain $\det(\boldsymbol{A})^{1/n} \leq \frac{1}{n}\text{Tr}[\boldsymbol{A}]$. We know already that the maximum volume ellipsoid is unique and we get a contradiction. It remains to show that $\mathcal{E}_\delta \subseteq K$ for small enough $\delta > 0$.

For a contact point $\boldsymbol{u} \in U$ we have $\langle (\boldsymbol{I}_n + \delta \boldsymbol{B})\boldsymbol{u}, \boldsymbol{u}\rangle = \|\boldsymbol{u}\|_2^2 + \delta \langle \boldsymbol{B}, \boldsymbol{u}\boldsymbol{u}^T\rangle > 1$ and in particular $\mathcal{E}_\delta$ does not even touch any of the contact points. One can also argue that every point $\boldsymbol{x} \in S^{n-1}$ that is very close to a contact point in $\|\cdot\|_2$ distance still has some slack (we skip the standard calculation here). Then it remains to consider $V := \{\boldsymbol{v} \in S^{n-1} \mid \text{dist}(\boldsymbol{v}, U) \geq \varepsilon\}$ where $\varepsilon := \frac{1}{2\|\boldsymbol{B}\|}$ suffices. Here $\text{dist}(\boldsymbol{v}, U) := \inf\{\|\boldsymbol{v} - \boldsymbol{u}\|_2 \mid \boldsymbol{u} \in U\}$ is the Euclidean distance. But V is a compact set with a positive distance from the boundary of K (the distance does not dependent on δ). Then it is clear that one can pick a small δ so that $\mathcal{E}_\delta \subseteq K$.



To get the bound of $m \leq \binom{n+1}{2}+1 \leq n^2+1$ note that we need to find the convex coefficients satisfying

$$\frac{\boldsymbol{I}_n}{n} \in \operatorname{conv}\{\boldsymbol{u}\boldsymbol{u}^T \mid \boldsymbol{u} \in U\}$$

But this is a system in dimension $\binom{n}{2} + n = \binom{n+1}{2}$ and the claim follows from Caratheodory's Theorem. □

We want to introduce a useful view on this result. Suppose that we take the weights $c_1, \dots, c_m > 0$ so that $\sum_{j=1}^m c_j \boldsymbol{x}_j \boldsymbol{x}_j^T = \boldsymbol{I}_n$. Let μ be a probability distribution that produces $\boldsymbol{x} \in \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ where $\Pr[\boldsymbol{x} = \boldsymbol{x}_j] = \frac{c_j}{n}$. Then $\mathbb{E}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}\boldsymbol{x}^T] = \frac{\boldsymbol{I}_n}{n}$. Such a distribution μ is called an *isotropic measure*.

For the sake of completeness we state the corresponding result for asymmetric bodies. Again we will be skipping the proof.

Theorem 2.5 (John)**.** *Let B_2^n be the maximum volume ellipsoid in a convex body K. Then there are contact points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ of K and B_2^n for $m \leq \binom{n+1}{2} + 1$ and scalars $c_1, \dots, c_m > 0$ so that*

$$\sum_{j=1}^m c_j \boldsymbol{x}_j = \boldsymbol{0} \quad \textit{and} \quad \boldsymbol{I}_n = \sum_{j=1}^m c_j \boldsymbol{x}_j \boldsymbol{x}_j^T \quad (*)$$

Note that the condition $\boldsymbol{I}_n = \sum_{j=1}^m c_j \boldsymbol{x}_j \boldsymbol{x}_j^T$ is invariant under flipping the sign of one of the $\boldsymbol{x}_j$'s. For a symmetric body K, $\boldsymbol{x}$ is a contact point iff $-\boldsymbol{x}$ is a contact point. But for asymmetric bodies it is clear that the contact points should not all be on one side of a hyperplane — otherwise one could move the center and scale the ball. Then $\sum_{j=1}^m c_j \boldsymbol{x}_j = \boldsymbol{0}$ is equivalent to saying that $\boldsymbol{0}$ is in the convex hull of the contact points.

We should also mention that the condition on the existence of the contact points is actually equivalent to B_2^n being the maximum volume ellipsoid. In other words, one can use the contact points to prove that K is already in John position. We just state the result due to Ball and skip the proof (which is somewhat reverse to the one we have seen above)

Theorem 2.6 (Ball)**.** *Let $K \supseteq B_2^n$ be a centrally symmetric convex body. Suppose there are $c_1, \dots, c_m > 0$ and points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \partial K \cap S^{n-1}$ with $\sum_{j=1}^m c_j \boldsymbol{x}_j \boldsymbol{x}_j^T = \boldsymbol{I}_n$. Then B_2^n is the maximum volume ellipsoid contained in K.*

2.3 Contact points and the Dvoretzky-Rogers Theorem

Suppose now that K is a symmetric convex body in John position. Then we know by John's Theorem that for any unit vector $\boldsymbol{x} \in S^{n-1}$ one has $\|\boldsymbol{x}\|_K \geq \frac{1}{\sqrt{n}}$. On the other hand we know that the set of contact points $C := \partial K \cap S^{n-1}$ is non-empty and one has $\|\boldsymbol{x}\|_K = 1$ for all those points $\boldsymbol{x} \in C$. The question that naturally arises is whether there is an *orthonormal basis* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ so that $\|\boldsymbol{x}_i\|_K \approx 1$. In fact, this is possible as we will see now. We begin by proving a simple lemma.

Lemma 2.7. *Let $K \supseteq B_2^n$ be a symmetric convex body in John's position. Then for any matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, there is a contact point $\boldsymbol{y} \in \partial K \cap B_2^n$ so that $\langle \boldsymbol{A}, \boldsymbol{y}\boldsymbol{y}^T \rangle \geq \frac{\mathrm{Tr}[\boldsymbol{A}]}{n}$.*

Proof. As K is in John's position, we know by Theorem 2.4 that there is an isotropic measure μ on the contact points so that $\mathbb{E}_{\boldsymbol{y}\sim\mu}[\boldsymbol{y}\boldsymbol{y}^T] = \frac{\boldsymbol{I}_n}{n}$. Then $\mathbb{E}_{\boldsymbol{y}\sim\mu}[\langle \boldsymbol{A}, \boldsymbol{y}\boldsymbol{y}^T \rangle] = \langle \boldsymbol{A}, \mathbb{E}_{\boldsymbol{y}\sim\mu}[\boldsymbol{y}\boldsymbol{y}^T] \rangle = \frac{\mathrm{Tr}[\boldsymbol{A}]}{n}$. Hence in particular there must be one contact point satisfying this inequality. $\square$

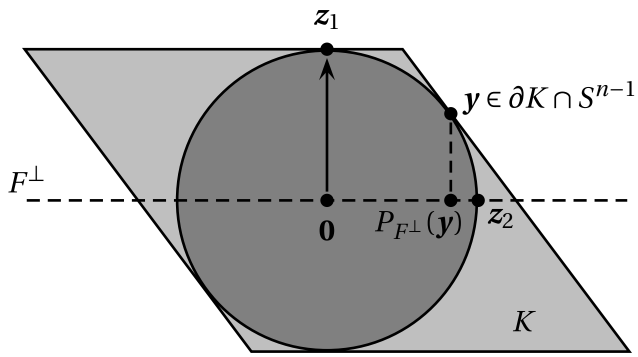
Now we will prove the Dvoretzky-Rogers Theorem which for a symmetric body K in John position picks an orthonormal basis $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ so that each $\|\boldsymbol{z}_k\|_K$ is lower bounded. We will pick the vectors iteratively one after the other. In iteration $k+1$ we will pick that contact point that has the largest projection on the orthogonal complement of $\mathrm{span}\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}$. Then we will see that $\|\boldsymbol{z}_k\|_K \geq \sqrt{1 - \frac{k-1}{n}}$. In particular the first couple of vectors have $\|\boldsymbol{z}_k\|_K \approx 1$ — later the guarantee deteriorates and for the last vector we can merely prove that $\|\boldsymbol{z}_n\|_K \geq \frac{1}{\sqrt{n}}$.

Theorem 2.8 (Dvoretzky-Rogers)**.** *Suppose that $K \subseteq \mathbb{R}^n$ is a symmetric convex body in John position. Then there exists an orthonormal basis $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ so that*

$$\sqrt{1 - \frac{k-1}{n}} \leq \|\boldsymbol{z}_k\|_K \leq \|\boldsymbol{z}_k\|_2 = 1 \quad \forall k = 1, \ldots, n$$

Proof. We select the vectors iteratively. Suppose we already picked $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$. Let $F := \mathrm{span}\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}$ and let $F^\perp$ be the subspace orthogonal to F. Define $P_{F^\perp} : \mathbb{R} \to F^\perp$ as the *orthogonal projection* onto $F^\perp$. In particular for $\boldsymbol{x} \in F$ one has $P_{F^\perp}(\boldsymbol{x}) = \boldsymbol{0}$ and for $\boldsymbol{x} \in F^\perp$ one has $P_{F^\perp}(\boldsymbol{x}) = \boldsymbol{x}$. Then considering this as a projection matrix one knows that $\mathrm{Tr}[\boldsymbol{P}_{F^\perp}] = \dim(F^\perp) \geq n - k$. We use the previous Lemma to find a contact point $\boldsymbol{y}$ satisfying

$$\|P_{F^\perp}(\boldsymbol{y})\|_2^2 \overset{\text{projection}}{=} \langle \boldsymbol{y}\boldsymbol{y}^T, \boldsymbol{P}_{F^\perp} \rangle \overset{\text{last lemma}}{\geq} \frac{\mathrm{Tr}[\boldsymbol{P}_{F^\perp}]}{n} \geq 1 - \frac{k}{n}$$



Example for $k = 1$

Then we normalize the projected contact point to $\boldsymbol{z}_{k+1} := \frac{P_{F^\perp}(\boldsymbol{y})}{\|P_{F^\perp}(\boldsymbol{y})\|_2}$ and estimate its $\|\cdot\|_K$-norm as

$$\|\boldsymbol{z}_{k+1}\|_K \cdot \underbrace{\|\boldsymbol{y}\|_{K^\circ}}_{=1} \stackrel{\text{Cauchy-Schwarz}}{\geq} \langle \boldsymbol{y}, \boldsymbol{z}_{k+1}\rangle \stackrel{(*)}{=} \|P_{F^\perp}(\boldsymbol{y})\|_2 \geq \sqrt{1 - \frac{k}{n}}$$

Here in $(*)$ we use that $\boldsymbol{z}_{k+1}$ is the vector $P_{F^\perp}(\boldsymbol{y})$ scaled to unit length. □

In fact, with a small trick one can even pick an orthonormal basis so that $\|\boldsymbol{b}_i\|_K = \Theta(1)$ for all $i = 1, \ldots, n$.

Lemma 2.9. *Let $K \supseteq B_2^n$ be a symmetric convex body in John position. Then there is an orthonormal basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ with $\frac{1}{4} \leq \|\boldsymbol{b}_i\|_K \leq 1$ for $i = 1, \ldots, n$.*

Proof. Consider again the sequence of orthogonal vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ from the previous lemma so that $\|\boldsymbol{z}_i\|_K \geq \frac{1}{\sqrt{2}}$ at least for $i = 1, \ldots, \frac{n}{2}$. In fact, we can sort the vectors so that $\|\boldsymbol{z}_1\|_K \geq \ldots \geq \|\boldsymbol{z}_n\|_K > 0$. Now consider a pair $(\boldsymbol{z}_i, \boldsymbol{z}_{n+1-i})$ where the 2nd vector is too short, say $\|\boldsymbol{z}_{n+1-i}\|_K < \frac{1}{4}$. Then we can "mix" the pair to a new pair of orthonormal vectors $\frac{1}{\sqrt{2}}(\boldsymbol{z}_i \pm \boldsymbol{z}_{n+1-i})$. Then the $\|\cdot\|_K$-norm can be lower and upper bounded by the triangle inequality

$$\Big\|\frac{1}{\sqrt{2}}\Big(\boldsymbol{z}_i \pm \boldsymbol{z}_{n+i-1}\Big)\Big\|_K \geq \frac{1}{\sqrt{2}}\Big(\frac{1}{\sqrt{2}} - \frac{1}{4}\Big) \geq \frac{1}{4}$$

Overall that proves the claim. □

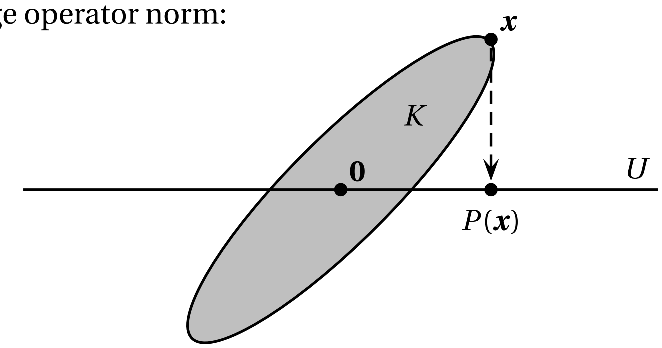
2.4 The Theorem of Kadets and Snobar

We will now see a beautiful application that can be derived from the existence of contact points. Consider the space $\mathbb{R}^n$ equipped with a norm $\|\cdot\|_K$ and let $U \subseteq \mathbb{R}^n$

be a subspace. We call a linear map $P : \mathbb{R}^n \to \mathbb{R}^n$ a *projection onto* U, if $P(\boldsymbol{x}) \in U$ and $P(P(\boldsymbol{x})) = P(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Equivalently, there is another subspace W with $W \cap U = \{\boldsymbol{0}\}$ and $\text{span}(U + W) = \mathbb{R}^n$ so that $P(\boldsymbol{u} + \boldsymbol{w}) = \boldsymbol{u}$ for all $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in w$. Such a linear map is an *operator* in the sense that it maps elements from the space $\mathbb{R}^n$ to the space $\mathbb{R}^n$. The *operator norm* of P with respect to the underlying norm $\|\cdot\|_K$ is the quantity

$$\|P\|_{\text{op}} := \sup\left\{\frac{\|P(\boldsymbol{x})\|_K}{\|\boldsymbol{x}\|_K} \mid \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}\right\}$$

In other words, $\|P\|_{\text{op}}$ gives the maximum "stretch" of any element $\boldsymbol{x}$ in terms the underlying norm and it is the minimum number so that $\|P(\boldsymbol{x})\|_K \leq \|P\|_{\text{op}} \cdot \|\boldsymbol{x}\|_K$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Observe that even in $n = 2$, the orthogonal projection can have an arbitrarily large operator norm:

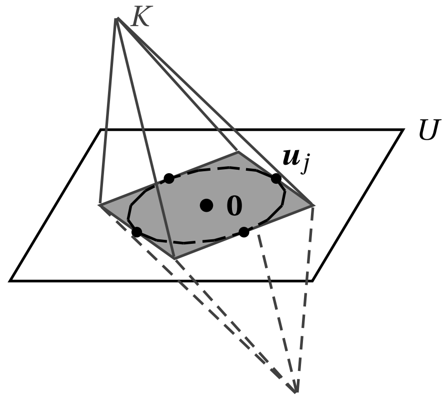


But it turns out that there always exists a projection P so that $\|P\|_{\text{op}}$ is even bounded by $\sqrt{k}$, where k is the dimension of the space U.

Theorem 2.10 (Kadets-Snobar)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $U \subseteq \mathbb{R}^n$ be a subspace with $k := \dim(U)$. Then there exists a projection $P : \mathbb{R}^n \to U$ with operator norm $\|P\|_{op} \leq \sqrt{k}$.*

Proof. Applying a linear transformation to K does not change whether or not such a transformation exists. So, suppose that $K \cap U$ is in John's position (with respect to the subspace U).

Then there are contact points $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \partial K \cap U$ of length $\|\boldsymbol{u}_i\|_2 = 1$ and coefficients $c_1, \ldots, c_m \geq 0$ with $\sum_{j=1}^m c_j = k$ so that $\sum_{j=1}^m c_j \boldsymbol{u}_j \boldsymbol{u}_j^T = \boldsymbol{I}_U$, where $\boldsymbol{I}_U = \begin{pmatrix} \boldsymbol{I}_k & 0 \\ 0 & 0 \end{pmatrix}$ is the identity matrix for the subspace U. Define the linear function $F_j : U \to \mathbb{R}$ with $F_j(\boldsymbol{x}) := \langle \boldsymbol{u}_j, \boldsymbol{x} \rangle$ and note that $F_j(\boldsymbol{x}) = \langle \boldsymbol{u}_j, \boldsymbol{x} \rangle \leq \|\boldsymbol{x}\|_K$ as $\|\boldsymbol{u}_j\|_{K^\circ} = 1$. Then we use the Hahn-Banach Theorem (Theorem 1.24) to extend this linear function to a linear function $\tilde{F}_j(\boldsymbol{x}) := \langle \tilde{\boldsymbol{u}}_j, \boldsymbol{x} \rangle$ with $\tilde{F}_j(\boldsymbol{x}) \leq \|\boldsymbol{x}\|_K$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Geometrically speaking, $\langle \tilde{\boldsymbol{u}}_j, \boldsymbol{x} \rangle \leq 1$ is a valid inequality for K going through the point $\boldsymbol{u}_j$.



Then we define a linear map

$$P(\boldsymbol{x}) := \sum_{j=1}^{m} c_j \langle \tilde{\boldsymbol{u}}_j, \boldsymbol{x} \rangle \boldsymbol{u}_j$$

Note that one has indeed $P : \mathbb{R}^n \to U$ and for a point $\boldsymbol{x} \in U$ one has

$$P(\boldsymbol{x}) = \underbrace{\sum_{j=1}^{m} c_j \underbrace{\langle \tilde{\boldsymbol{u}}_j, \boldsymbol{x} \rangle}_{=\langle \boldsymbol{u}_j, \boldsymbol{x} \rangle} \boldsymbol{u}_j}_{=\boldsymbol{I}_U \boldsymbol{x}} = \boldsymbol{x}.$$

In particular P is a projection onto U. Next, let $\boldsymbol{x}^* \in K$ be the point attaining the operator norm and in the following calculation we denote $\boldsymbol{y}^* \in (K \cap U)^\circ$ as the point that is dual to $P(\boldsymbol{x}^*)$ in the sense that $\|P(\boldsymbol{x}^*)\|_K = \langle P(\boldsymbol{x}^*), \boldsymbol{y}^* \rangle$, see Lemma 1.7. Then

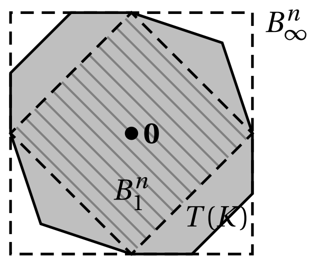
$$\begin{aligned}
\|P\|_{\mathrm{op}} &= \sup_{\|\boldsymbol{x}\|_K = 1} \|P(\boldsymbol{x})\|_K = \|P(\boldsymbol{x}^*)\|_K = \langle P(\boldsymbol{x}^*), \boldsymbol{y}^* \rangle \\
&\overset{\text{Def } P}{=} \sum_{j=1}^{m} c_j \underbrace{\langle \tilde{\boldsymbol{u}}_j, \boldsymbol{x}^* \rangle}_{|\cdot| \le 1} \cdot \langle \boldsymbol{u}_j, \boldsymbol{y}^* \rangle \\
&\le \sum_{j=1}^{m} c_j \cdot |\langle \boldsymbol{u}_j, \boldsymbol{y}^* \rangle| = \sum_{j=1}^{m} \sqrt{c_j} \cdot \sqrt{c_j} \cdot |\langle \boldsymbol{u}_j, \boldsymbol{y}^* \rangle| \\
&\overset{\text{Cauchy-Schwarz}}{\le} \Big(\underbrace{\sum_{j=1}^{m} c_j}_{=k} \Big)^{1/2} \cdot \Big(\underbrace{\sum_{j=1}^{m} c_j \langle \boldsymbol{u}_j, \boldsymbol{y}^* \rangle^2}_{=\|\boldsymbol{y}^*\|_2^2 \le 1} \Big)^{1/2} \le \sqrt{k}
\end{aligned}$$

where $\|\boldsymbol{y}^*\|_2 \le 1$ follows from $(K \cap U)^\circ \subseteq B_2^n \cap U$. This shows the claim. $\square$

2.5 Auerbach's Lemma

We want to state and prove a result due to Auerbach that sandwiches a symmetric convex body between B_1^n and B_∞^n rather than between B_2^n and $\sqrt{n}B_2^n$. Note that the proof is more related to the one of Lewis' Lemma that we will see later in Section 6.2.

Lemma 2.11 (Auerbach's Lemma — Geometric version)**.** *For any symmetric convex body $K \subseteq \mathbb{R}^n$, there is a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $B_1^n \subseteq T(K) \subseteq B_\infty^n$.*



Proof. Let $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n \in K$ be points maximizing $|\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)|$ (those exist due to compactness). Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation so that $T(\boldsymbol{a}_i) = \boldsymbol{e}_i$ and let $\tilde{K} := T(K)$ be the transformed body. Then $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n \in \tilde{K}$ and hence $B_1^n \subseteq \tilde{K}$ by convexity. Also, we can see that $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n \in \tilde{K}$ still form an optimum solution to the optimization problem $\max\{|\det(\tilde{\boldsymbol{a}}_1,\dots,\tilde{\boldsymbol{a}}_n)| : \tilde{\boldsymbol{a}}_1,\dots,\tilde{\boldsymbol{a}}_n \in \tilde{K}\}$ (using that $\det(\boldsymbol{TA}) = \det(\boldsymbol{T})\cdot\det(\boldsymbol{A})$). Then by optimality we know that $\|\boldsymbol{e}_i\|_{\tilde{K}} = 1$ for all $i \in [n]$. It remains to prove the following:
Claim. *One has $\tilde{K} \subseteq B_\infty^n$.*
Proof of Claim. Suppose for the sake of contradiction that there is a point $\boldsymbol{x} \in \tilde{K}$ with $\langle \boldsymbol{e}_i, \boldsymbol{x}\rangle > 1$ (the case $\langle \boldsymbol{e}_i, \boldsymbol{x}\rangle < -1$ is analogous). Then the vector $\boldsymbol{d} := \boldsymbol{x} - \boldsymbol{e}_i$ has $d_i > 0$ and $\boldsymbol{e}_i + \delta\boldsymbol{d} \in \tilde{K}$ for all $0 \le \delta \le 1$ by convexity. Let $\boldsymbol{D} := (\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{d},\boldsymbol{0},\dots,\boldsymbol{0}) \in \mathbb{R}^{n\times n}$ be the matrix that has the vector $\boldsymbol{d}$ as ith column and all other columns are $\boldsymbol{0}$. Then replacing $\boldsymbol{e}_i$ by $\boldsymbol{e}_i + \delta\boldsymbol{d}$ increases the objective function by

$$\det(\boldsymbol{I}_n + \delta\boldsymbol{D}) - \det(\boldsymbol{I}_n) \overset{\text{up to } \delta^2 \text{ terms}}{\approx} \delta\mathrm{Tr}[\boldsymbol{D}] = \delta d_i > 0$$

Hence for $\delta > 0$ small enough, we get a contradiction to optimality. $\square$

The result of Auerbach has also a functional-analytic form that can be easily derived from the geometric statement above:

Lemma 2.12 (Auerbach's Lemma)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then there are bases $\{\boldsymbol{a}_i\}_{i\in[n]}$ and $\{\boldsymbol{b}_i\}_{i\in[n]}$ of $\mathbb{R}^n$ so that $\|\boldsymbol{a}_i\|_K = 1$ and $\|\boldsymbol{b}_i\|_{K^\circ} = 1$ with $\langle \boldsymbol{a}_i, \boldsymbol{b}_i\rangle = 1$ for all $i \in [n]$ and $\langle \boldsymbol{a}_i, \boldsymbol{b}_j\rangle = 0$ for all $i \neq j$.*

Proof. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be the linear map so that $B_1^n \subseteq A(K) \subseteq B_\infty^n$. Let $\boldsymbol{a}_i$ be the ith column of $\boldsymbol{A}^{-1}$ and let $\boldsymbol{b}_i$ be the ith row of $\boldsymbol{A}$. Then $\|\boldsymbol{a}_i\|_K = \|(\boldsymbol{A}^{-1})^i\|_K = \|\boldsymbol{A}^{-1}\boldsymbol{e}_i\|_K = \|\boldsymbol{e}_i\|_{A(K)} = 1$. Moreover $\|\boldsymbol{b}_i\|_{K^\circ} = \|\boldsymbol{A}^T\boldsymbol{e}_i\|_{K^\circ} = \|\boldsymbol{e}_i\|_{(\boldsymbol{A}^T)^{-1}(K^\circ)} = \|\boldsymbol{e}_i\|_{A(K)^\circ} = 1$. Also for any indices $i,j \in [n]$ one has $\langle \boldsymbol{b}_i, \boldsymbol{a}_j\rangle = \langle \boldsymbol{A}_i, (\boldsymbol{A}^{-1})^j\rangle = (\boldsymbol{A}\boldsymbol{A}^{-1})_{ij} = (\boldsymbol{I}_n)_{ij}$ as claimed. $\square$

Note that none of the bases $\{\boldsymbol{a}_i\}_{i\in[n]}$ or $\{\boldsymbol{b}_i\}_{i\in[n]}$ has to necessarily be orthogonal. The basis $\{\boldsymbol{a}_i\}_{i\in[n]}$ is called an *Auerbach basis* for the normed vector space $(\mathbb{R}^n, \|\cdot\|_K)$ (note that $\{\boldsymbol{a}_i\}_{i\in[n]}$ uniquely determines the choice of $\{\boldsymbol{b}_i\}_{i\in[n]}$).

2.6 Exercises

Exercise 2.1.
Let $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| \leq 1 \ \forall i \in [N]\}$ be a symmetric polytope and suppose $\mathcal{E}$ is the maximum volume ellipsoid contained in K. Show that there is a subset of indices $I \subseteq [N]$ with $|I| \leq n^2+1$ so that $\mathcal{E}$ is still the largest volume ellipsoid in $Q := \{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| \leq 1 \ \forall i \in I\}$.

Exercise 2.2.
For a matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ we define a quantity called γ_2*-norm* as $\gamma_2(\boldsymbol{M}) := \inf\{\max\{\|\boldsymbol{A}_i\|_2 \cdot \|\boldsymbol{B}^j\|_2 : i,j\} \mid \boldsymbol{M} = \boldsymbol{A}\boldsymbol{B} \text{ with } \boldsymbol{A} \in \mathbb{R}^{m\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times n} \text{ for some } k\}$ where $\boldsymbol{A}_i$ is the ith row of $\boldsymbol{A}$ and $\boldsymbol{B}^j$ is the jth column of $\boldsymbol{B}$. Prove that for any matrix $\boldsymbol{M} \in [-1,1]^{m\times n}$ with $k := \text{rank}(\boldsymbol{M})$ one has $\gamma_2(\boldsymbol{M}) \leq \sqrt{k}$.
Hint. Consider the rank factorization $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{B}$. Then apply a linear transformation so that $\text{conv}\{\pm\boldsymbol{A}_i : i \in [m]\}$ is in John position. After this transformation how long can the vectors $\boldsymbol{A}_i$ and $\boldsymbol{B}^j$ be?

Exercise 2.3.
Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $\mathcal{E}$ be a *minimum* volume ellipsoid with $K \subseteq \mathcal{E}$.

(i) Prove that $\frac{1}{\sqrt{n}}\mathcal{E} \subseteq K$.
(ii) Show that there are points $S \subseteq \text{vert}(K)$ (meaning extreme points of K) with $|S| \leq n^2+1$ so that $\mathcal{E}$ is also the minimum volume ellipsoid with $\text{conv}((-S)\cup S) \subseteq \mathcal{E}$.

Hint. Use polarity!

Exercise 2.4.
In this exercise we want to elaborate on an application that is due to Naor (2011).

(i) Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body with $B_2^n \subseteq K$ and suppose there are $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in S^{n-1} \cap \partial K$ with $\boldsymbol{I}_n \preceq \sum_{i=1}^m c_i \boldsymbol{u}_i \boldsymbol{u}_i^T \preceq (1+\varepsilon)\boldsymbol{I}_n$ for $\varepsilon \geq 0$. Prove that $K \subseteq \sqrt{(1+\varepsilon)n} B_2^n$.

A remarkable spectral sparsification result of Batson, Spielman, Srivastava (2008) says the following (which you may use without a proof): *For any $\varepsilon > 0$ and any vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$ with $\sum_{i=1}^m \boldsymbol{v}_i \boldsymbol{v}_i^T = \boldsymbol{I}_n$ there are coefficients $\boldsymbol{s} \in \mathbb{R}^m_{\geq 0}$ with $|supp(\boldsymbol{s})| \leq O(n/\varepsilon^2)$ so that $\boldsymbol{I}_n \preceq \sum_{i=1}^m s_i \boldsymbol{v}_i \boldsymbol{v}_i^T \preceq (1+\varepsilon)\boldsymbol{I}_n$*. Use this to prove the following.

(ii) Let $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| \leq 1 \ \forall i \in [N]\}$ be a symmetric polytope that is in John position. Then there are indices $I \subseteq [N]$ with $|I| \leq O(n/\varepsilon^2)$ so that $Q = \{\boldsymbol{x} \in \mathbb{R}^n \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| \leq 1 \ \forall i \in I\}$ satisfies $B_2^n \subseteq Q \subseteq \sqrt{(1+\varepsilon)n} B_2^n$.

Exercise 2.5.

Suppose that $K \subseteq \mathbb{R}^n$ is a convex body with the property that (i) $\|(\sigma_1 x_1, \ldots, \sigma_n x_n)\|_K = \|\boldsymbol{x}\|_K$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and all $\boldsymbol{\sigma} \in \{-1,1\}^n$ and $\|(x_{\pi(1)}, \ldots, x_{\pi(n)})\|_K = \|\boldsymbol{x}\|_K$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and every permutation $\pi : [n] \to [n]$. Prove that for some $r > 0$, the body rK is in John position.

Chapter 3

Isoperimetric inequalities and concentration of measure

Concentration of measure is a phenomenon that is tremendously useful in convex geometry as well as other areas such as combinatorics. Here we will give several inequalities in different settings. We begin with a concentration result that has a nice geometric proof.

3.1 Concentration on the sphere

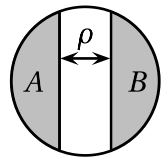
Recall that $S^{n-1} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 = 1\}$ is the *sphere* and σ is the uniform measure on S^{n-1}. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *L-Lipschitz* if $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_2$ for all $\boldsymbol{x}, \boldsymbol{y}$. In case that $f : S^{n-1} \to \mathbb{R}$ is 1-Lipschitz, we will be able to prove that $\Pr_{\boldsymbol{x} \in S^{n-1}}[|f(\boldsymbol{x}) - \mu| \leq t] \leq \exp(-\Theta(t^2 n))$ where μ is the *median* (and in other settings we will be using the mean). Luckily, there is no need to handle the function explicitly. Simply define $A := \{\boldsymbol{x} \in S^{n-1} \mid f(\boldsymbol{x}) \leq \mu\}$ — which by definition is a set with measure $\sigma(A) \geq \frac{1}{2}$ — and consider the set $A_t := \{\boldsymbol{x} \in S^{n-1} \mid d(\boldsymbol{x}, A) \leq t\}$, where $d(\boldsymbol{x}, A) := \inf\{\|\boldsymbol{y} - \boldsymbol{x}\|_2 : y \in A\}$ is the Euclidean distance to A. Then the concentration inequality immediately follows from the pure geometric statement of $\sigma(A_t) \geq 1 - \exp(-\Theta(t^2 n))$. In other words, we need to show that most of the measure on a sphere cannot be too far from a set of measure $1/2$.

We begin with a simple geometric argument for the ball:

Lemma 3.1. *Let μ be the uniform measure on the ball B_2^n and let $d(A, B)$ be the distance between A and B. Then for $A, B \subseteq B_2^n$, then*

$$\min\{\mu(A), \mu(B)\} \leq \exp\left(-\frac{n}{8} \cdot d(A, B)^2\right)$$

Proof. Let us abbreviate $\alpha := \min\{\mu(A), \mu(B)\}$ and let $\rho := d(A,B)$ be the distance.



Then for any $\boldsymbol{a} \in A, \boldsymbol{b} \in B$ one has

$$\|\boldsymbol{a}+\boldsymbol{b}\|_2^2 = 2\underbrace{\|\boldsymbol{a}\|_2^2}_{\leq 1} + 2\underbrace{\|\boldsymbol{b}\|_2^2}_{\leq 1} - \underbrace{\|\boldsymbol{a}-\boldsymbol{b}\|_2^2}_{\geq \rho^2} \leq 4-\rho^2$$

using the parallelogram law. That means $\frac{A+B}{2} \subseteq \sqrt{1-\frac{1}{4}\rho^2} \cdot B_2^n$. Hence

$$\alpha \overset{\text{Brunn-Minkowski}}{\leq} \mu\Big(\frac{A+B}{2}\Big) \leq \Big(\sqrt{1-\frac{\rho^2}{4}}\Big)^n \leq \exp\Big(-n \cdot \frac{\rho^2}{8}\Big).$$

using $\sqrt{1-x} \leq e^{-x/2}$ for $0 \leq x \leq 1$. □

One can also rearrange the statement of Lemma 3.1 to

$$d(A,B) \leq \sqrt{\frac{8}{n} \cdot \ln\Big(\frac{1}{\min\{\mu(A),\mu(B)\}}\Big)}$$

For example, if $A, B \subseteq S^{n-1}$ with $\mu(A), \mu(B) \geq \Omega(1)$, then $d(A,B) \leq O(\frac{1}{\sqrt{n}})$.

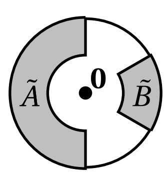
We will now prove a measure concentration for the sphere:

Theorem 3.2. *Let $A \subseteq S^{n-1}$ with $\sigma(A) = \frac{1}{2}$ and abbreviate $A_t := \{\boldsymbol{x} \in S^{n-1} : d(\boldsymbol{x}, A) \leq t\}$. Then for $t > 0$ one has*

$$\sigma(A_t) \geq 1 - 2\exp(-\Theta(t^2 n)).$$

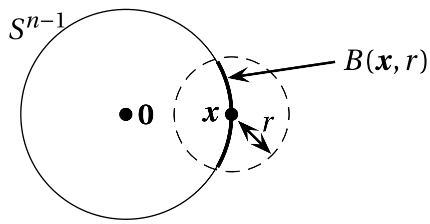
Proof. Let $B := S^{n-1} \setminus A_t$ be the points in the sphere that have distance bigger than t. Define

$$\tilde{A} := \Big\{\lambda A \mid \frac{1}{2} \leq \lambda \leq 1\Big\} \quad \text{and} \quad \tilde{B} := \Big\{\lambda B \mid \frac{1}{2} \leq \lambda \leq 1\Big\}$$



Then the distance of those sets is $d(\tilde{A},\tilde{B}) \geq \frac{t}{2}$. By the previous Lemma 3.1 we get the inequality in $(1-\frac{1}{2^n})\sigma(B) = \mu(\tilde{B}) \leq \exp(-\Theta(t^2 n))$. Then we have our upper bound on $\sigma(B)$. $\square$

For $\boldsymbol{x} \in S^{n-1}$ we define $B(\boldsymbol{x}, r) := \{\boldsymbol{y} \in S^{n-1} \mid \|\boldsymbol{x}-\boldsymbol{y}\|_2 \leq r\}$ as the *geodesic ball of radius* r (also called a *spherical cap*).



It is known that spherical caps minimizes the "measure expansion". We state the result without a proof:

Theorem 3.3 (Lévy, Schmidt)**.** *For any set $A \subseteq S^{n-1}$, take a geodesic ball $B(\boldsymbol{x}, r) \subseteq S^{n-1}$ with r chosen so that $\sigma(A) = \mu(B(\boldsymbol{x}, r))$. Then $\sigma(A_t) \geq \sigma(B(\boldsymbol{x}, r+t))$.*

For a function $f : S^{n-1} \to \mathbb{R}$ we abbreviate $\text{mean}(f) := \mathbb{E}[f(\boldsymbol{x})]$ and $\text{median}(f)$ denotes any median of $f(\boldsymbol{x})$ under the distribution $\boldsymbol{x} \sim S^{n-1}$. We state two more results with explicit constants without detailed proof:

Theorem 3.4. *Let $f : S^{n-1} \to \mathbb{R}$ be an L-Lipschitz function. Then for any $t \geq 0$, $\Pr_{\boldsymbol{x}\sim S^{n-1}}[|f(\boldsymbol{x}) - \text{median}(f)| > t \cdot L] \leq 4e^{-nt^2/2}$.*

See for example [Mat02] for a derivation.

Theorem 3.5. *Let $f : S^{n-1} \to \mathbb{R}$ be an L-Lipschitz function. Then for any $t \geq 0$, $\Pr_{\boldsymbol{x}\sim S^{n-1}}[|f(\boldsymbol{x}) - \text{mean}(f)| > t \cdot L] \leq 64e^{-nt^2/64}$.*

Proof. After scaling assume $L = 1$. Then using concentration one can show that $|\text{median}(f) - \text{mean}(f)| \leq \frac{12}{\sqrt{n}}$ (see exercises of [Mat02]). First note that the claim is vacuous if $t \leq \frac{16}{\sqrt{n}}$ as the right hand side is bigger than 1. Then for $t \geq \frac{16}{\sqrt{n}}$ we get

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim S^{n-1}}[|f(\boldsymbol{x}) - \text{mean}(f)| > t] &\leq \Pr_{\boldsymbol{x}\sim S^{n-1}}\Big[|f(\boldsymbol{x}) - \text{median}(f)| > t - \frac{12}{\sqrt{n}}\Big] \\ &\leq 4\exp\Big(-n\cdot\Big(t-\frac{12}{\sqrt{n}}\Big)^2/2\Big) \leq 64\exp(-nt^2/64)\end{aligned}$$

as one can verify. $\square$

3.2 Isoperimetric inequality in Gaussian space

Next, we want to prove a concentration result for Gaussians. It turns out that such a result has a rather short proof using the Prékopa-Leindler Inequality. We take a small detour in order to spell out where the "magic" in the proof comes from. For the sake of simplicity let us set $\lambda := \frac{1}{2}$ from here on.

Recall that the Prékopa-Leindler inequality says that for functions $f, g, h : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $h(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}) \geq f(\boldsymbol{x})^{1/2} g(\boldsymbol{y})^{1/2}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ one obtains a lower bound on the integral of $\int_{\mathbb{R}^n} h(\boldsymbol{x}) d\boldsymbol{x} \geq (\int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x})^{1/2} (\int_{\mathbb{R}^n} g(\boldsymbol{x}) d\boldsymbol{x})^{1/2}$. Now we are interested in such an inequality for Gaussian space in the form of $\mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[h(\boldsymbol{x})] \geq \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x})]^{1/2} \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[g(\boldsymbol{x})]^{1/2}$. Observe that the Gaussian density has the property that the density $\gamma_n(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y})$ is a lot higher than the product of densities $\gamma_n(\boldsymbol{x})^{1/2} \cdot \gamma_n(\boldsymbol{y})^{1/2}$ if $\boldsymbol{x}$ and $\boldsymbol{y}$ are far apart. One can use this to build in a "discount factor" in the assumptions.

Lemma 3.6 (Prékopa-Leindler for Gaussian Space)**.** *Let $f, g, h : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be measurable functions with*

$$h\Big(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\Big) \geq \exp\Big(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{8}\Big) \cdot f(\boldsymbol{x})^{1/2} g(\boldsymbol{y})^{1/2} \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Then

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[h(\boldsymbol{x})] \geq \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x})]^{1/2} \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[g(\boldsymbol{x})]^{1/2}.$$

Proof. We will apply the original Prékopa-Leindler inequality for the functions

$$\tilde{f}(\boldsymbol{x}) = \gamma_n(\boldsymbol{x}) \cdot f(\boldsymbol{x}) \quad \text{and} \quad \tilde{g}(\boldsymbol{x}) = \gamma_n(\boldsymbol{x}) \cdot g(\boldsymbol{x}) \quad \text{and} \quad \tilde{h}(\boldsymbol{x}) = \gamma_n(\boldsymbol{x}) \cdot h(\boldsymbol{x})$$

Then we can verify that

$$\begin{aligned}
\tilde{h}\Big(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\Big) &= \frac{1}{(2\pi)^{n/2}} \exp\Big(-\frac{1}{2}\Big\|\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\Big\|_2^2\Big) \cdot h\Big(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\Big) \\
&= \frac{1}{(2\pi)^{n/2}} \big(e^{-\frac{1}{2}\|\boldsymbol{x}\|_2^2}\big)^{1/2} \big(e^{-\frac{1}{2}\|\boldsymbol{y}\|_2^2}\big)^{1/2} \exp\Big(\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{8}\Big) \cdot h\Big(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\Big) \\
&\overset{\text{assumption}}{\geq} \Big(\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|\boldsymbol{x}\|_2^2} \cdot f(\boldsymbol{x})\Big)^{1/2} \Big(\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|\boldsymbol{y}\|_2^2} \cdot g(\boldsymbol{y})\Big)^{1/2} \\
&= \tilde{f}(\boldsymbol{x})^{1/2} \tilde{g}(\boldsymbol{x})^{1/2}.
\end{aligned}$$

This means we can indeed apply Prékopa-Leindler to the functions $\tilde{f}, \tilde{g}, \tilde{h}$ and

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[h(\boldsymbol{x})] &= \int_{\mathbb{R}^n} \tilde{h}(\boldsymbol{x}) d\boldsymbol{x} \overset{\text{PL}}{\geq} \Big(\int_{\mathbb{R}^n} \tilde{f}(\boldsymbol{x}) d\boldsymbol{x}\Big)^{1/2} \Big(\int_{\mathbb{R}^n} \tilde{g}(\boldsymbol{x}) d\boldsymbol{x}\Big)^{1/2} \\
&= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x})]^{1/2} \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[g(\boldsymbol{x})]^{1/2}
\end{aligned}$$

and the lemma is proven. □

Now we can prove that the Gaussian measure expands quickly:

Theorem 3.7. *Let $A \subseteq \mathbb{R}^n$ be a non-empty measurable set. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[e^{d(\boldsymbol{x},A)^2/4}\right] \leq \frac{1}{\gamma_n(A)}.$$

Proof. Define functions

$$f(\boldsymbol{x}) := \exp(d(\boldsymbol{x},A)^2/4) \quad \text{and} \quad g(\boldsymbol{x}) := \mathbf{1}_A(\boldsymbol{x}) \quad \text{and} \quad h(\boldsymbol{x}) := 1$$

Then we verify that the assumption of Lemma 3.6 is satisfied. For $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$ one has

$$f(\boldsymbol{x})^{1/2}\cdot g(\boldsymbol{y})^{1/2} e^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/8} = \underbrace{e^{d(\boldsymbol{x},A)^2/8}\cdot \mathbf{1}_A(\boldsymbol{y})}_{\leq e^{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/8}} \cdot e^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/8} \leq 1 = h\Big(\frac{1}{2}\boldsymbol{x}+\frac{1}{2}\boldsymbol{y}\Big)$$

Here we use that for $\boldsymbol{y} \notin A$ we have $\mathbf{1}_A(\boldsymbol{y}) = 0$ and there is nothing to show, while for $\boldsymbol{y} \in A$, we have the upper bound of $d(\boldsymbol{x},A) \leq \|\boldsymbol{x}-\boldsymbol{y}\|_2$. Then by the Prékopa-Leindler inequality for Gaussian space we get

$$1 = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[h(\boldsymbol{x})]^2 \stackrel{\text{Lem 3.6}}{\geq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x})]\cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[g(\boldsymbol{x})] = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\exp(d(\boldsymbol{x},A)^2/4)\right]\cdot \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[\mathbf{1}_A(\boldsymbol{x})]}_{=\gamma_n(A)}$$

and rearranging gives the claim. □

It is actually a rather standard approach in concentration to first obtain an upper bound on $\mathbb{E}[\exp(\text{distance of } \boldsymbol{x} \text{ to } A)]$ and then derive that almost all points $\boldsymbol{x}$ are not too far from A.

Lemma 3.8. *Let $A \subseteq \mathbb{R}^n$ be a set with $\gamma_n(A) = \frac{1}{2}$, then $\gamma_n(A_t) \geq 1 - 2\exp(-t^2/4)$.*

Proof. Simply write

$$\Pr_{\boldsymbol{x}\sim\gamma_n}[d(\boldsymbol{x},A) \geq t] = \Pr_{\boldsymbol{x}\sim\gamma_n}\left[e^{d(\boldsymbol{x},A)^2/4} \geq e^{t^2/4}\right] \stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{d(\boldsymbol{x},A)^2/4}]}{e^{t^2/4}} \stackrel{\text{Thm 3.7}}{\leq} 2e^{-t^2/4}$$

□

We also state the beautiful fact that among all sets of identical Gaussian measure, a halfspace minimizes the expansion. Again we omit a proof:

Theorem 3.9 (Gaussian Isoperimetric Inequality - Borell, Sudakov-Tsirelson)**.** *Let $A \subseteq \mathbb{R}^n$ be a measurable set and let $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \leq \lambda\}$ be a halfspace with $\gamma_n(A) = \gamma_n(H)$. Then for any $t \geq 0$ one has $\gamma_n(A_t) \geq \gamma_n(H_t)$.*

It is often convinient to state concentration with respect to mean instead of median. Note that in an exercise we will see that for every 1-Lipschitz function these differ by at most a constant and one could derive the following claim from that (possibly with worse constants). Hence we will skip the proof.

Theorem 3.10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be an L-Lipschitz function. Then for any $s \geq 0$ one has $\Pr_{\boldsymbol{x} \sim \gamma_n}[|f(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{y} \sim \gamma_n}[f(\boldsymbol{y})]| \geq s \cdot L] \leq 2\exp(-s^2/4)$.*

3.3 Talagrand's inequality

For this section, we follow the exposition from the book *The Probabilistic Method* of Alon & Spencer [AS16]. Suppose that $\Omega = \Omega_1 \times \ldots \times \Omega_n$ is a product space with a product measure μ, meaning that $\boldsymbol{x} \sim \mu$ is a random vector from Ω and every coordinate is drawn independently from some distribution. Talagrand's inequality gives a concentration inequality for *every* possible product distribution; in particular the sets Ω_i do not need to come from $\mathbb{R}$. This generality also comes at a cost — the inequality is a bit hard to parse and to understand. For vectors $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, let

$$\text{unequal}(\boldsymbol{x}, \boldsymbol{y})_i := \begin{cases} 1 & \text{if } x_i \neq y_i \\ 0 & \text{if } x_i = y_i \end{cases}$$

Then for $\boldsymbol{x} \in \Omega$ one defines $U_A(\boldsymbol{x}) := \{\boldsymbol{s} \in \{0,1\}^n : \boldsymbol{s} \geq \text{unequal}(\boldsymbol{x}, \boldsymbol{y}) \text{ for some } \boldsymbol{y} \in A\}$ and

$$\phi_A(\boldsymbol{x}) := \min\{\|\boldsymbol{s}\|_2 \mid \boldsymbol{s} \in \text{conv}(U_A(\boldsymbol{x}))\}$$

Note that $\phi_A(\boldsymbol{x})$ is a distance function that is also called the *convex distance* in the literature. For the sake of illustration fix an $\boldsymbol{x} \in \Omega$ and fix the vector $\boldsymbol{s} \in \text{conv}(U_A(\boldsymbol{x}))$ attaining the distance. Then the definition provides the existance of a distribution ν over vectors in A so that for every coordinate $i \in [n]$ one has $\Pr_{\boldsymbol{y} \sim \nu}[x_i \neq y_i] \leq s_i$.

Theorem 3.11. *Let $\Omega = \Omega_1 \times \ldots \times \Omega_n$ be a product space with a product distribution μ and let $A \subseteq \Omega$ with $\mu(A) > 0$. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu}\left[\exp(\phi_A(\boldsymbol{x})^2/4)\right] \leq \frac{1}{\mu_n(A)}$$

Proof. We prove the claim by induction over the dimension n. For $n = 1$ we have

$$\mathop{\mathbb{E}}_{x\sim\mu}\left[\exp\left(\frac{1}{4}\phi_A(x)^2\right)\right] = e^0 \cdot \mu(A) + (1-\mu(A)) \cdot e^{1/4} \le \frac{1}{\mu(A)}$$

as one can easily check.

Now we consider an $(n+1)$-dim vector $(\boldsymbol{x},\omega) \in (\Omega_1 \times \ldots \times \Omega_n) \times \Omega_{n+1}$. To simplify notation we denote μ as the product measure in the appropriate dimension. Let

$$A_\omega := \{\boldsymbol{x} \in \Omega_1 \times \ldots \times \Omega_n \mid (\boldsymbol{x},\omega) \in A\}$$

be the *slices* of A for a given last coordinate. Moreover, let

$$B := \{\boldsymbol{x} \in \Omega_1 \times \ldots \times \Omega_n \mid \exists \omega \in \Omega_{n+1} : (\boldsymbol{x},\omega) \in A\}$$

be the set that can be reached by changing the last coordinate.

Observe that in order to move from a random point $(\boldsymbol{x},\omega)$ to $\mathrm{conv}(U_A(\boldsymbol{x},\boldsymbol{\omega}))$ we have two options: (A) change the last coordinate or (B) leave the last coordinate unchanged. We use the fact that

$$\begin{array}{lccc} (A) & \boldsymbol{s} \in \mathrm{conv}(U_B(\boldsymbol{x})) & \Rightarrow & (\boldsymbol{s},1) \in \mathrm{conv}(U_A(\boldsymbol{x},\omega)) \\ (B) & \boldsymbol{t} \in \mathrm{conv}(U_{A_\omega}(\boldsymbol{x})) & \Rightarrow & (\boldsymbol{t},0) \in \mathrm{conv}(U_A(\boldsymbol{x},\omega)) \end{array}$$

Pick the points $\boldsymbol{s} \in \mathrm{conv}(U_B(\boldsymbol{x}))$ and $\boldsymbol{t} \in \mathrm{conv}(A_\omega(\boldsymbol{x}))$ minimizing the length $\|\cdot\|_2$. Note that for (A) we use the monotonicity in the definition of $U_A(\boldsymbol{x},\omega)$ as the point $\boldsymbol{s} \in \mathrm{conv}(U_B(\boldsymbol{x}))$ may have a component from A_ω.

Then for any $0 \le \lambda \le 1$ we have

$$\lambda \begin{pmatrix} \boldsymbol{s} \\ 1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} \boldsymbol{t} \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda \boldsymbol{s} + (1-\lambda)\boldsymbol{t} \\ \lambda \end{pmatrix} \in \mathrm{conv}(U_A(\boldsymbol{x},\omega))$$

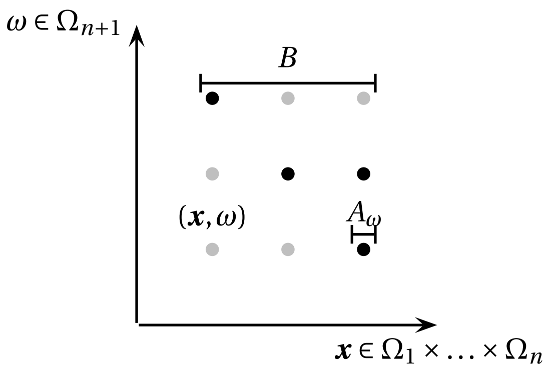


Fig 1: View in $\Omega_1 \times \ldots \times \Omega_{n+1}$

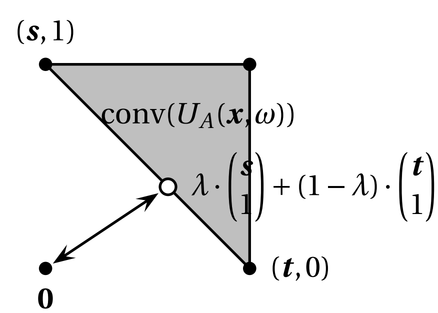


Fig 2: View in $\{0,1\}^{n+1}$

Then we can upper bound the distance for a point by

$$\begin{aligned}
\phi_A(\boldsymbol{x},\omega)^2 &\leq \left\| \begin{pmatrix} \lambda \boldsymbol{s} + (1-\lambda)\boldsymbol{t} \\ \lambda \end{pmatrix} \right\|_2^2 \overset{\text{Phytagoras}}{=} \|\lambda \boldsymbol{s} + (1-\lambda)\boldsymbol{t}\|_2^2 + \lambda^2 \quad (*) \\
&\overset{\text{convexity}}{\leq} \lambda \|\boldsymbol{s}\|_2^2 + (1-\lambda)\|\boldsymbol{t}\|_2^2 + \lambda^2 \overset{\text{choice of } \boldsymbol{s},\boldsymbol{t}}{=} \lambda \cdot \phi_B(\boldsymbol{x})^2 + (1-\lambda)\cdot \phi_{A_\omega}(\boldsymbol{x})^2 + \lambda^2
\end{aligned}$$

We use the inequality $(*)$ and average over the sampled point. Here it will be crucial that we may choose λ dependent on the outcome of ω.

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\boldsymbol{x},\omega)}\left[\exp\left(\frac{1}{4}\phi_A(\boldsymbol{x},\omega)^2\right)\right] &\overset{(*)}{\leq} \mathop{\mathbb{E}}_{\omega}\left[e^{\lambda^2/4}\mathop{\mathbb{E}}_{\boldsymbol{x}}\left[\exp\left(\frac{1}{4}\phi_B(\boldsymbol{x})^2\right)^{\lambda}\exp\left(\frac{1}{4}\phi_{A_\omega}(\boldsymbol{x})^2\right)^{1-\lambda}\right]\right] \\
&\overset{\text{Hölder}}{\leq} \mathop{\mathbb{E}}_{\omega}\left[e^{\lambda^2/4}\mathop{\mathbb{E}}_{\boldsymbol{x}}\left[\exp\left(\frac{1}{4}\phi_B(\boldsymbol{x})^2\right)\right]^{\lambda}\cdot\mathop{\mathbb{E}}_{\boldsymbol{x}}\left[\exp\left(\frac{1}{4}\phi_{A_\omega}(\boldsymbol{x})^2\right)\right]^{1-\lambda}\right] \\
&\overset{\text{induction}}{\leq} \mathop{\mathbb{E}}_{\omega}\left[e^{\lambda^2/4}\left(\frac{1}{\mu(A_\omega)}\right)^{1-\lambda}\left(\frac{1}{\mu(B)}\right)^{\lambda}\right] \\
&\overset{(**)}{\leq} \mathop{\mathbb{E}}_{\omega}\left[\frac{1}{\mu(B)}\cdot\left(2-\frac{\mu(A_\omega)}{\mu(B)}\right)\right] \\
&= \frac{1}{\mu(B)}\cdot\left(2-\frac{\mu(A)}{\mu(B)}\right) = \frac{1}{\mu(A)}\cdot\underbrace{\frac{\mu(A)}{\mu(B)}\cdot\left(2-\frac{\mu(A)}{\mu(B)}\right)}_{\leq 1 \text{ as } 0\leq\frac{\mu(A)}{\mu(B)}\leq 1} \leq \frac{1}{\mu(A)}
\end{aligned}$$

Here we can justify $(**)$ as follows. For a fixed ω, set $r := \frac{\mu(A_\omega)}{\mu(B)} \in [0,1]$ as the ratio of the measures. Then

$$e^{\lambda^2/4}\left(\frac{1}{\mu(A_\omega)}\right)^{1-\lambda}\left(\frac{1}{\mu(B)}\right)^{\lambda} = \frac{1}{\mu(B)}\cdot e^{\lambda^2/4}r^{-(1-\lambda)} \leq \frac{1}{\mu(B)}\cdot(2-r)$$

if we choose

$$\lambda(r) := \begin{cases} 2\ln(\frac{1}{r}) & \text{if } e^{-1/2} \leq r \leq 1 \\ 0 & 0 \leq r < e^{-1/2} \end{cases}$$

This finishes the proof. $\square$

If the product spaces is $\{-1,1\}^n$ we can simplify the statement of Talagrand's Theorem.

Corollary 3.12 (Talagrand on Hypercube I)**.** *Let μ be a product measure on $\{-1,1\}^n$ and let $A \subseteq \{-1,1\}^n$. Then $\mathbb{E}_{\boldsymbol{x}\sim\mu}\left[\exp(\frac{1}{16}d(\boldsymbol{x},conv(A))^2)\right] \leq \frac{1}{\mu(A)}$.*

Note that the extra factor of $\frac{1}{4}$ comes from scaling $\{0,1\}^n$ to $\{-1,1\}^n$ which doubles each distance.

Corollary 3.13 (Talagrand on Hypercube II)**.** *Let μ be a product measure on $\{-1,1\}^n$ and let $A \subseteq \{-1,1\}^n$. Then for $t \geq 0$ one has* $\Pr_{\boldsymbol{x}\sim\mu}\left[d(\boldsymbol{x}, \mathrm{conv}(A)) \geq t\right] \leq \frac{\exp(-t^2/16)}{\mu(A)}$.

Proof. Similar to earlier proofs we bound

$$\Pr_{\boldsymbol{x}\sim\mu}\left[d(\boldsymbol{x}, \mathrm{conv}(A)) \geq t\right] = \Pr\left[e^{d(\boldsymbol{x},A)^2/16} \geq e^{t^2/16}\right] \leq \frac{\mathbb{E}\left[e^{d(\boldsymbol{x},\mathrm{conv}(A))^2/16}\right]}{e^{t^2/16}} \leq \frac{\exp(-t^2/16)}{\mu(A)}$$

□

An application of Talagrand's Inequality

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and μ be a measure. A number median(f) is called a *median* of f if

$$\Pr_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x}) \geq \mathrm{median}(f)] \geq \frac{1}{2} \quad \text{and} \quad \Pr_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x}) \leq \mathrm{median}(f)] \geq \frac{1}{2}$$

Note that the median does not have to be unique.

We can also give one application that uses Talagrand's Concentration inequality:

Theorem 3.14. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex 1-Lipschitz function. Then*

$$\Pr_{\boldsymbol{x}\sim\{-1,1\}^n}[|f(\boldsymbol{x}) - median(f)| \geq t] \leq 4e^{-t^2/16}$$

Proof. Let μ_n be the uniform measure on $\{-1,1\}^n$. For $s \in \mathbb{R}$, we define $A_s := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \leq s\}$. Note that the sets A_s are convex. We prove the following:
Claim. *For any s and $t \geq 0$ one has* $\mu_n(A_{s+t}) \geq 1 - \frac{1}{\mu_n(A_s)} e^{-t^2/16}$.

Proof of claim. We have $\Pr_{\boldsymbol{x}\in\{-1,1\}^n}[d(\boldsymbol{x}, \mathrm{conv}(A_s \cap \{-1,1\}^n)) \leq t] \geq 1 - \frac{1}{\mu_n(A_s)} e^{-t^2/16}$ by Talagrand's inequality (Cor 3.13). So, fix an $\boldsymbol{x} \in \{-1,1\}^n$ satisfying this event. By definition of the distance function, there is a $\boldsymbol{y} \in \mathrm{conv}(A_s \cap \{-1,1\}^n) \subseteq A_s$ so that $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq t$. By the Lipschitz property $f(\boldsymbol{x}) \leq f(\boldsymbol{y}) + \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq s + t$. □

We will need the claim twice. Let us assume for the sake of a simpler notation that indeed $\mu_n(A_{\mathrm{median}(f)}) = \frac{1}{2}$. First we can get that

$$\mu_n(A_{\mathrm{median}(f)+t}) \overset{\text{Claim}}{\geq} 1 - \frac{1}{\mu_n(A_{\mathrm{median}(f)})} \cdot e^{-t^2/16} = 1 - 2e^{-t^2/16}.$$

Next,

$$\frac{1}{2} = \mu_n(A_{\mathrm{median}(f)}) \overset{\text{Claim}}{\geq} 1 - \frac{1}{\mu_n(A_{\mathrm{median}(f)-t})} e^{-t^2/16}$$

which can be rearranged to $\mu_n(A_{\mathrm{median}(f)-t}) \leq 2e^{-t^2/16}$. It follows that $\mu_n(A_{\mathrm{median}(f)+t} \setminus A_{\mathrm{median}(f)-t}) \geq 1 - 4e^{-t^2/16}$. □

3.4 The subgaussian norm

Frequently we will deal with random variables that are not necessarily Gaussian but that have tails that decay at least as quickly as this is the case for Gaussians. We will now discuss properties of such random variables in detail. For this section we follow the exposition in Vershynin [Ver19]. Here we will focus on *mean-zero* random variables as this suffices for our purposes (and we do not need to tediously point out which condition works also for non-centered random variables and which one does not). First we want to elaborate what exactly we mean by "tail bound". Luckily, many possible conditions for tail bounds are actually equivalent.

Lemma 3.15 (Conditions of Sub-Gaussian tails)**.** *Let $X \in \mathbb{R}$ be a random variable with $\mathbb{E}[X] = 0$. The following statements are equivalent in the sense that if condition i holds with $s_i > 0$ then there is an $s_j \in [\frac{s_i}{C}, Cs_i]$ so that also condition j holds where $C > 0$ is a universal constant.*

- *Condition 1: One has $\Pr[|X| \geq t] \leq 2\exp(-t^2/s_1^2)$ for all $t \geq 0$.*
- *Condition 2: One has $\mathbb{E}[|X|^p]^{1/p} \leq s_2\sqrt{p}$ for all $p \geq 1$.*
- *Condition 3: One has $\mathbb{E}[\exp(X^2/s_3^2)] \leq 2$.*
- *Condition 4: One has $\mathbb{E}[\exp(\lambda X)] \leq \exp(s_4^2\lambda^2)$ for all $\lambda \in \mathbb{R}$.*

The proof is a bit lengthy (though not difficult) so we skip it here. See [Ver19] for details. So we could pick pretty much any parameter s_i and use it to quantify the tail bounds of a random variable. We pick Condition 3:

Definition 3.16. Let $X \in \mathbb{R}$ be a random variable. We define the *sub-gaussian norm* as[1] as

$$\|X\|_{\psi_2} := \inf\left\{s > 0 : \mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right] \leq 2\right\}$$

Note that for $X \sim N(0,\sigma^2)$ we indeed have $\|X\|_{\psi_2} = \Theta(\sigma)$ as expected. We summarize a few useful properties — in particular the subgaussian norm is indeed a norm on the set of mean-zero random variables.

Lemma 3.17. *In the following let X as well as $X_1,\dots,X_N$ be jointly distributed mean-zero random variables.*

(i) One has $\mathbb{E}[\max\{|X_1|,\dots,|X_N|\}] \leq O(\sqrt{\log(N)}) \cdot \max\{\|X_i\|_{\psi_2} : i \in [N]\}$

[1] The notation comes from the more general concept of an *Orlicz norm*.

(ii) *One has* $\|tX\|_{\psi_2} = |t| \cdot \|X\|_{\psi_2}$ *for all* $t \in \mathbb{R}$.

(iii) *One has* $\|X_1 + X_2\|_{\psi_2} \leq \|X_1\|_{\psi_2} + \|X_2\|_{\psi_2}$ *(even if* X_1, X_2 *are dependent).*

(iv) *If* $X_1, \ldots, X_N$ *are independent then* $\|X_1 + \ldots + X_N\|_{\psi_2} \leq C \cdot (\sum_{i=1}^N \|X_i\|_{\psi_2}^2)^{1/2}$ *for a universal constant* $C > 0$.

Proof. We only prove (i) and (iv). For (i), we set $\sigma := \max\{\|X_i\|_{\psi_2} : i = 1, \ldots, N\}$. Then some constant $C_0 > 0$ and all $s \geq 1$ one has

$$\Pr\left[\exists i \in [N] : |X_i| \geq C_0 s\sigma\sqrt{\log(N)}\right] \leq N \cdot e^{-2s^2 \log(N)} = N^{1-2s^2}$$

Then

$$\begin{aligned}
\mathbb{E}\left[\max|X_i|\right] &\leq C_0\sigma\sqrt{\log(N)} + \sum_{s \in \mathbb{Z}_{\geq 1}} (1+s)C_0\sigma\sqrt{\log(N)} \cdot \Pr\left[\exists i \in [N] : X_i \geq C_0 s\sigma\sqrt{\log(N)}\right] \\
&\leq C_0\sigma\sqrt{\log(N)} \cdot \Big(1 + \underbrace{\sum_{s \in \mathbb{Z}_{\geq 1}} (1+s)N^{1-2s^2}}_{\leq \text{constant}}\Big) \leq C_1\sigma\sqrt{\log(N)}
\end{aligned}$$

for some constant $C_1 > 0$. Note that this upper bound indeed does not require independence[2].

For (iv) we abbreviate $\sigma_i := \Theta(\|X_i\|_{\psi_2})$ as the value so that for $\lambda \in \mathbb{R}$ one has $\mathbb{E}[e^{\lambda X_i}] \leq e^{\sigma_i^2 \lambda^2}$. Then one has

$$\mathbb{E}\left[\exp\Big(\lambda \sum_{i=1}^N X_i\Big)\right] = \prod_{i=1}^N \mathbb{E}[\exp(\lambda X_i)] \leq \prod_{i=1}^N \exp(\sigma_i^2\lambda^2) = \exp(\|\boldsymbol{\sigma}\|_2^2\lambda^2)$$

which gives that $\|X_1 + \ldots + X_n\|_{\psi_2} \leq O(\|\boldsymbol{\sigma}\|_2)$. $\square$

3.5 Khintchine's Inequality

Consider independent random variables $x_1, \ldots, x_n \in \{-1, 1\}$ with $\Pr[x_i = +1] = \Pr[x_1 = -1] = \frac{1}{2}$ for all i. Such random variables are also called *Rademacher random variables.* Now, take a vector $\boldsymbol{a} \in \mathbb{R}^n$ — say for the sake of simplicity normalized so that $\|\boldsymbol{a}\|_2 = 1$ — and consider the outcome of the sum $\langle \boldsymbol{a}, \boldsymbol{x} \rangle := \sum_{i=1}^n a_i x_i$.

[2] A 2nd popular way of deriving the same bound is as follows: Let $\boldsymbol{Y} := (X_1, \ldots, X_N) \in \mathbb{R}^N$ and $p := \log_2(N)$ and recall the useful estimate that $\|\boldsymbol{Y}\|_p \leq \|\boldsymbol{Y}\|_\infty \leq 2\|\boldsymbol{Y}\|_p$. Then $\mathbb{E}[\|\boldsymbol{Y}\|_\infty] \leq 2\mathbb{E}[\|\boldsymbol{Y}\|_p] = 2\mathbb{E}[(\sum_{i=1}^N |X_i|^p)^{1/p}] \leq 2(\sum_{i=1}^N \mathbb{E}[|X_i|^p])^{1/p} \leq 2(N \cdot C_2\sqrt{p})^{1/p} \leq C_3\sqrt{\log(N)}$ using Lemma 3.15.(II) and Hölders Inequality in the form $\mathbb{E}[Z^{1/p}] = \mathbb{E}[1^{1-1/p}Z^{1/p}] \leq \mathbb{E}[1]^{1-1/p}\mathbb{E}[Z]^{1/p} = \mathbb{E}[Z]^{1/p}$ for any non-negative random variable Z.

Clearly $\mathbb{E}[\langle \boldsymbol{a}, \boldsymbol{x}\rangle] = 0$ and $\mathrm{Var}[\langle \boldsymbol{a}, \boldsymbol{x}\rangle] = 1$. We are wondering for fixed $p > 0$, how the *pth moment* $\mathbb{E}[|\langle \boldsymbol{a}, \boldsymbol{x}\rangle|^p]^{1/p}$ of that random variable is going to behave. If $p = \Theta(1)$, then also that moment will be constant. But if p is large, this puts higher weight on outliers and if the a_i's are tiny we might get a deviation as we know it from a Gaussian. On the other hand, if $a_1 = 1$ and $a_i = 0$ for $i \neq 1$, then the pth moment is just 1. In fact, the Inequality of Khintchine provides us with upper bounds and lower bounds on $\mathbb{E}[|\langle \boldsymbol{a}, \boldsymbol{x}\rangle|^p]^{1/p}$. This is a useful tool at numerous places.

Theorem 3.18 (Khintchine)**.** *For $n \geq 1$ and $p > 0$ and $\boldsymbol{a} \in \mathbb{R}^n$ one has*

$$\left\{ \begin{array}{ll} 1 & \text{if } p \geq 2 \\ C' & \text{if } 0 < p \leq 2 \end{array} \right\} \cdot \|\boldsymbol{a}\|_2 \leq \mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\left[|\langle \boldsymbol{a}, \boldsymbol{x}\rangle|^p\right]^{1/p} \leq \|\boldsymbol{a}\|_2 \cdot \left\{ \begin{array}{ll} C\sqrt{p} & \text{if } p \geq 2 \\ 1 & \text{if } 0 < p \leq 2 \end{array} \right\}$$

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

Proof. W.l.o.g. one can scale the coefficients so that $\|\boldsymbol{a}\|_2 = 1$ and consider the sum $X := \sum_{i=1}^n a_i x_i$. Note that $\mathbb{E}[X] = \mathbb{E}[\langle \boldsymbol{a}, \boldsymbol{x}\rangle] = 0$ and $\mathbb{E}[X^2] = \mathbb{E}[\langle \boldsymbol{a}, \boldsymbol{x}\rangle^2] = 1$ and X satisfies Gaussian tail bounds. That will be all we need. We distinguish several cases.
Upperbound for $p \geq 2$. By Lemma 3.17.(ii)+(iv) we have $\|X\|_{\psi_2} \leq C_1(\sum_{i=1}^n a_i^2 \|x_i\|_{\psi_2}^2)^{1/2} \leq C_2\|\boldsymbol{a}\|_2^2 = C_2$ for some constants $C_1, C_2 > 0$. We conclude by Lemma 3.15 that $\mathbb{E}[|X|^p]^{1/p} \leq C\sqrt{p}$ for some other constant $C > 0$.
Lower bound for $p \geq 2$. Then we use Jensen's inequality with the fact that $x \mapsto x^{p/2}$ is convex as $p/2 \geq 1$ to get $\mathbb{E}[|X|^p] = \mathbb{E}[(|X|^2)^{p/2}] \geq (\mathbb{E}[|X|]^2)^{p/2} = 1$.
Upper Bound for $0 < p \leq 2$. Now the function $x \mapsto x^{p/2}$ is concave for $0 \leq x < \infty$ and so $\mathbb{E}[|X|^p] = \mathbb{E}[(X^2)^{p/2}] \leq \mathbb{E}[X^2]^{p/2} = 1$ by using again Jensen's inequality.
Lower bound for $1 \leq p \leq 2$. We use Hölder's inequality to obtain

$$1 = \mathbb{E}[X^2] = \mathbb{E}\left[|X|^{2/3} \cdot (|X|^4)^{1/3}\right] \leq \mathbb{E}\left[|X|\right]^{2/3} \cdot \mathbb{E}\left[|X|^4\right]^{1/3} \leq \mathbb{E}\left[|X|\right]^{2/3} \cdot ((C\sqrt{4})^4)^{1/3}$$

This can be rearranged to obtain $\mathbb{E}[|X|^p] \geq \mathbb{E}[|X|] \geq C' > 0$ for some universal constant.
Lower bound for $0 < p < 1$. Similar to the last case. ◻

3.6 Kahane's Inequality

Suppose we consider $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$ and we consider the random vector $\boldsymbol{X} := \sum_{i=1}^n x_i \boldsymbol{a}_i$ where $\boldsymbol{x} \sim \{-1,1\}^n$ is drawn uniformly at random. Then what can we say about the value of $\mathbb{E}[\|\boldsymbol{X}\|_K^p]^{1/p}$ compared to $\mathbb{E}[\|\boldsymbol{X}\|_K]$, where $\|\cdot\|_K$ is an arbitrary norm? Clearly, for $p \geq 1$, an extra factor of $O(\sqrt{p})$ will be needed even if the $\boldsymbol{a}_i$'s are identical. And this turns out that this is the worst case.

Theorem 3.19 (Kahane)**.** *Let $K \subseteq \mathbb{R}^m$ be a symmetric convex body. Then for $p \geq 1$ and $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n \in \mathbb{R}^m$ one has*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\Big[\Big\|\sum_{i=1}^n x_i \boldsymbol{a}_i\Big\|_K^p\Big]^{1/p} \leq O(\sqrt{p}) \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\Big[\Big\|\sum_{i=1}^n x_i \boldsymbol{a}_i\Big\|_K\Big]$$

Proof. We abbreviate $\boldsymbol{X} := \sum_{i=1}^n x_i \boldsymbol{a}_i$ as the produced random vector. Then the quantity $\sigma^2 := \max\{\sum_{i=1}^n \langle \boldsymbol{a}_i, \boldsymbol{b}\rangle^2 \mid \boldsymbol{b} \in K^\circ\}$ will be a good proxy for the variance of $\|\boldsymbol{X}\|_K$. Note that σ^2 would be in particular large if the $\boldsymbol{a}_i$ are co-linear.
Claim I. *The function $f(\boldsymbol{x}) := \|\sum_{i=1}^n x_i \boldsymbol{a}_i\|_K$ is convex and σ-Lipschitz.*
Proof of Claim. Convexity follows from $\|\cdot\|_K$ being a norm. Next, take vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and let $\boldsymbol{z}$ be the *dual element* to $\sum_{i=1}^n y_i \boldsymbol{a}_i$, i.e. $\|\sum_{i=1}^n y_i \boldsymbol{a}_i\|_K = \langle \sum_{i=1}^n y_i \boldsymbol{a}_i, \boldsymbol{z}\rangle$ with $\|\boldsymbol{z}\|_{K^\circ} = 1$ (see again Lemma 1.7). Then

$$\begin{aligned}
|f(\boldsymbol{x}+\boldsymbol{y}) - f(\boldsymbol{x})| &= \Big|\Big\|\sum_{i=1}^n (x_i + y_i)\boldsymbol{a}_i\Big\|_K - \Big\|\sum_{i=1}^n x_i \boldsymbol{a}_i\Big\|_K\Big| \overset{\substack{\text{Reverse}\\\text{triangle}}}{\leq} \Big\|\sum_{i=1}^n y_i \boldsymbol{a}_i\Big\|_K \\
&\leq \Big|\langle \sum_{i=1}^n y_i \boldsymbol{a}_i, \boldsymbol{z}\rangle\Big| \leq \sum_{i=1}^n |y_i| \cdot |\langle \boldsymbol{a}_i, \boldsymbol{z}\rangle| \overset{\text{Cauchy-S.}}{\leq} \|\boldsymbol{y}\|_2 \cdot \underbrace{\Big(\sum_{i=1}^n \langle \boldsymbol{a}_i, \boldsymbol{z}\rangle^2\Big)^{1/2}}_{\leq \sigma} \quad \square
\end{aligned}$$

We will also need an upperbound on the quantity σ:
Claim II. *One has $\sigma \leq C \cdot \mathbb{E}[\|\boldsymbol{X}\|_K]$ for some constant $C > 0$.*
Let $\boldsymbol{b} \in K^\circ$ be the element attaining the maximum that defines σ. Then

$$\begin{aligned}
\sigma &= \Big(\sum_{i=1}^n \langle \boldsymbol{a}_i, \boldsymbol{b}\rangle^2\Big)^{1/2} \overset{\text{Khintchine}}{\leq} C \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\Big[\Big|\sum_{i=1}^n x_i \langle \boldsymbol{a}_i, \boldsymbol{b}\rangle\Big|\Big] \\
&= C \cdot \mathbb{E}[|\langle \boldsymbol{X}, \boldsymbol{b}\rangle|] \overset{\text{Cauchy-S.}}{\leq} C \cdot \mathbb{E}\big[\|\boldsymbol{X}\|_K \cdot \underbrace{\|\boldsymbol{b}\|_{K^\circ}}_{=1}\big] \quad \square
\end{aligned}$$

The function f is convex and σ-Lipschitz. Moreover by Markov's inequality we can get the rather crude bound of $\text{median}(f) \leq 2\mathbb{E}[\|\boldsymbol{X}\|_K]$. Then by Talagrand's inequality (see Theorem 3.14) we obtain concentration of the form $\Pr[|\|\boldsymbol{X}\|_K - \text{median}(f)| > t\sigma] \leq 4e^{-t^2/16}$. Then using a calculation analogous to Khintchine's inequality we obtain

$$\mathbb{E}\big[\|\boldsymbol{X}\|_K^p\big]^{1/p} \leq \text{median}(f) + O(\sqrt{p} \cdot \sigma) \overset{\text{Claim II}}{\leq} O(\sqrt{p}) \cdot \mathbb{E}[\|\boldsymbol{X}\|_K]$$

That shows the claim. $\square$

3.7 Log-concave measures

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *log-concave* if

$$f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y}) \geq f(\boldsymbol{x})^{1-\lambda} f(\boldsymbol{y})^{\lambda} \quad \forall 0 < \lambda < 1 \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Similarly a probability measure μ on $\mathbb{R}^n$ is called *log-concave* if

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

for all $0 < \lambda < 1$ and all measurable sets $A, B \subseteq \mathbb{R}^n$. In fact, these notions are connected:

Lemma 3.20. *If the density function $\mu(\boldsymbol{x})$ of a measure is log-concave, then also the measure μ itself is log-concave.*

The proof is a straightforward application of the Prékopa-Leindler Inequality. Again, we have somewhat abused notation and used the same symbol μ for the measure and the density of the measure. One consequence is that the Gaussian measure γ_n is log-concave. Moreover any convex body induces a natural log-concave measure:

Lemma 3.21. *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then the measure μ defined by $\mu(A) := \frac{Vol_n(A\cap K)}{Vol_n(K)}$ is log-concave.*

Proof. By Brunn-Minkowski III (Theorem 1.41) we have $\mathrm{Vol}_n(\lambda \cdot (A \cap K) + (1-\lambda) \cdot (B \cap K)) \geq \mathrm{Vol}_n(A \cap K)^{\lambda} \cdot \mathrm{Vol}_n(B \cap K)^{1-\lambda}$. □

We have observed earlier that log-concave functions have an exponential decay. That leads to the suspicion that log-concave measures satisfy some form of concentration. And indeed we can prove a rather general result (where we are a bit loose with constants in order to simplify the exposition):

Lemma 3.22 (Borell)**.** *Let μ be a log-concave measure and let $A \subseteq \mathbb{R}^n$ be a symmetric convex set with $\mu(A) \geq \frac{3}{4}$. Then $\mu(t \cdot A) \geq 1 - 2^{-t/2}$ for $t \geq 4$.*

Proof. First observe that

$$\left(\frac{2}{t} \cdot (\mathbb{R}^n \setminus tA) + \left(1 - \frac{2}{t}\right) \cdot A\right) = \Big((\mathbb{R}^n \setminus 2A) + \underbrace{\left(1 - \frac{2}{t}\right)}_{<1} A)\Big) \subseteq \mathbb{R}^n \setminus A \qquad (*)$$

using the Reverse Triangle Inequality and that $\|\cdot\|_A$ is a norm. Then

$$\frac{1}{4} \geq \mu(\mathbb{R}^n \setminus A) \overset{(*)}{\geq} \mu\left(\frac{2}{t}(\mathbb{R}^n \setminus tA) + \left(1 - \frac{2}{t}\right) \cdot A\right) \overset{\text{log concavity}}{\geq} \mu(\mathbb{R}^n \setminus tA)^{2/t} \underbrace{\mu(A)^{1-2/t}}_{\geq (3/4)^{1/2}}$$

which can be rearranged to $\mu(\mathbb{R}^n \setminus tA) \leq (\frac{1}{4} \cdot (\frac{4}{3})^{1/2})^{t/2} \leq (1/2)^{t/2}$. $\square$

3.8 Exercises

Exercise 3.1.
Find a 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ so that $\Pr_{\boldsymbol{x}\sim\{-1,1\}^n}[|f(\boldsymbol{x}) - \text{median}(f)| \geq c_1 n^{1/4}] \geq c_2$ where $c_1, c_2 > 0$ are constants, $n \in \mathbb{N}$ is arbitrary and median(f) denotes the median w.r.t. the distribution $\boldsymbol{x} \sim \{-1,1\}^n$.
Hint. You may use the following fact without proof: For some small enough constants $c', c'' > 0$ one has $\Pr_{\boldsymbol{x}\sim\{-1,1\}^n}[\sum_{i=1}^n x_i \geq c'\sqrt{n}] \geq c''$.
Remark. This exercise shows that the convexity assumption in Talagrand's Theorem cannot be dropped.

Exercise 3.2.
Let $X : \Omega \to \mathbb{R}$ be a random variable with the property $\Pr[|X - \text{median}(X)| \geq t] \leq c_1 \exp(-c_2 t^2)$ for all $t \geq 0$ where $c_1, c_2 > 0$ are constants. Prove that $|\text{median}(X) - \mathbb{E}[X]| \leq c_3$ where $c_3 := c_3(c_1, c_2)$ is a constant only dependent on c_1 and c_2.

Exercise 3.3.
Let $H \subseteq \mathbb{R}^n$ be a subspace with $k := \dim(H)$ and let $d(\boldsymbol{x}, H) := \min\{\|\boldsymbol{x} - \boldsymbol{y}\|_2 : \boldsymbol{y} \in H\}$ be the Euclidean distance to H.

(i) Prove that $\mathbb{E}_{\boldsymbol{x}\sim\{-1,1\}^n}[d(\boldsymbol{x}, H)^2] = n - k$.

(ii) Prove that $\mathbb{E}_{\boldsymbol{x}\sim\{-1,1\}^n}[d(\boldsymbol{x}, H)] \leq \sqrt{n-k}$.

(iii) Prove that $\Pr_{\boldsymbol{x}\sim\{-1,1\}^n}[|d(\boldsymbol{x}, H) - \sqrt{n-k}| \geq t] \leq c_1 \exp(-c_2 t^2)$ for some universal constants $c_1, c_2 > 0$ and any $t > 0$.

Hint. You may use without proof the fact that for a $\mathbb{R}$-valued random variable X one has $|\text{median}(X) - \text{mean}(X)| \leq C\|X\|_{\psi_2}$ for some constant $C > 0$. You may also use without proof that for an $\mathbb{R}$-valued random variable X with $\mathbb{E}[X] = 0$ and any values $u \in \mathbb{R}$, $s > 0$ and $D_1 > 0$ one has

$$\Big(\Pr[|X - u| \geq t] \leq D_1 \exp(-t^2/s^2) \quad \forall t \geq 0\Big) \quad \Longrightarrow \quad \|X\|_{\psi_2} \leq D_2 \cdot s$$

where $D_2 > 0$ is a constant that only depends on D_1.

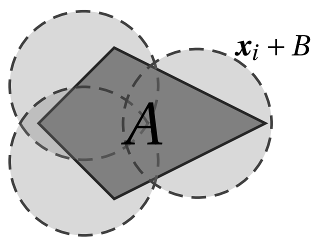
Exercise 3.4.
Let μ be a log concave measure on $\mathbb{R}$ and let $X \sim \mu$ be a random variable distributed according to μ.

1. Prove that there is a $C := C(\mu) > 0$ so that $\Pr[|X| \geq t] \leq 2^{-t/2}$ for all $t \geq C$.
2. Find a log concave measure μ so that indeed $\Pr[|X| \geq t] \geq 2^{-c_1 t}$ for all $t \geq c_2$ for some constants $c_1, c_2 > 0$.

Chapter 4

Covering numbers

For two convex bodies $A, B \subseteq \mathbb{R}^n$ we define the *covering number* $N(A, B)$ as the minimum number of translates of B necessary to cover A. In other words, the covering number is the minimum number N so that there are points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^n$ with $A \subseteq \bigcup_{i=1}^{N} (\boldsymbol{x}_i + B)$.



There is a natural relation between covering numbers and the volume of the involved sets. In particular it is a simple observation that $N(A, B) \geq \mathrm{Vol}_n(A)/\mathrm{Vol}_n(B)$ and this inequality could only be tight if A could be *partitioned* into disjoint copies of B. But most of the time the *shapes* of A and B do not allow such an efficient covering. For example one can have $\mathrm{Vol}_n(A) = \mathrm{Vol}_n(B)$, but still $N(A, B)$ can be arbitrarily large if A and B are "long and skinny" in very different directions. Nevertheless, one of the main insights of this chapter will be that for any two convex bodies $A, B \subseteq \mathbb{R}^n$ one has

$$4^{-n} \frac{\mathrm{Vol}_n(A - B)}{\mathrm{Vol}_n(B)} \leq N(A, B) \leq 4^n \frac{\mathrm{Vol}_n(A - B)}{\mathrm{Vol}_n(B)}$$

This is a surprisingly tight inequality showing that the relative shapes of A and B matter very little.

4.1 A few basic results in covering numbers

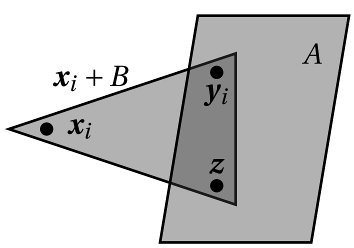
We start discussing a couple of basic, but useful facts. Note that in the definition of the covering number $N(A,B)$, the centers $\boldsymbol{x}_i$ might not be inside A. Hence, a variant is $\bar{N}(A,B) := \min\{N \in \mathbb{N} \mid \exists \boldsymbol{x}_1,\dots,\boldsymbol{x}_N \in A : A \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + B)\}$ where the centers of the translations have to lie inside of A. For some bodies B one can have $\bar{N}(A,B) > N(A,B)$, but for example if B is a Euclidean ball the numbers are identical.

There are a couple of facts:

Lemma 4.1. *The following holds:*

(1) For convex bodies $A, B \subseteq \mathbb{R}^n$ one has $\bar{N}(A, B-B) \leq N(A,B) \leq \bar{N}(A,B)$.
(2) For a convex body A and $r > 0$ one has $N(A, rB_2^n) = \bar{N}(A, rB_2^n)$.
(3) For convex bodies $A, B \subseteq \mathbb{R}^n$ and an invertible linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ one has $N(A,B) = N(T(A), T(B))$.
(4) For convex bodies A, B, C one has $N(A,B) \leq N(A,C) \cdot N(C,B)$.
(5) For convex bodies $A, B \subseteq \mathbb{R}^n$ one has $N(A, (A-A) \cap (B-B)) \leq N(A,B)$.

Proof. The claims (1)-(4) are very straightforward to show. But we will prove (5), which an interesting claim as the set $(A-A) \cap (B-B)$ is a symmetric convex set that is potentially a lot smaller than B. W.l.o.g. assume that $\mathbf{0} \in A$ and $\mathbf{0} \in B$ so that $A - A \supseteq A$ and $B - B \supseteq B$. So, suppose that $A \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + B)$. For a fixed index i pick an element $\boldsymbol{y}_i \in (\boldsymbol{x}_i + B) \cap A$ (which should exist, otherwise that translate was redundant). For $\boldsymbol{z} \in A \cap (\boldsymbol{x}_i + B)$ one can now see that $\|\boldsymbol{z} - \boldsymbol{y}_i\|_{A-A} \leq 1$ as well as $\|\boldsymbol{z} - \boldsymbol{y}_i\|_{B-B} \leq 1$. Then $A \subseteq \bigcup_{i=1}^N (\boldsymbol{y}_i + ((A-A) \cap (B-B)))$.

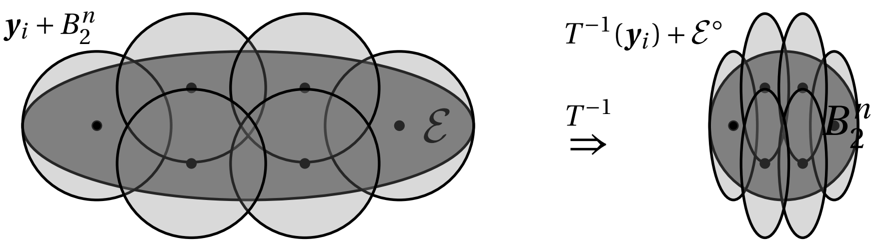


□

Next, we will see that for ellipsoids, the reverse covering number equals the covering number of the polars:

Lemma 4.2. *For ellipsoids $\mathcal{E}_1, \mathcal{E}_2$ one has $N(\mathcal{E}_1, \mathcal{E}_2) = N(\mathcal{E}_2^\circ, \mathcal{E}_1^\circ)$.*

Proof. After a linear transformation that does not affect the statement, we may assume that $\mathcal{E}_2 = B_2^n$. In other words, it suffices to prove that $N(\mathcal{E}, B_2^n) = N(B_2^n, \mathcal{E}^\circ)$ for an ellipsoid $\mathcal{E}$. Let T be the invertible linear map with $\mathcal{E} = T(B_2^n)$ and choose orthonormal vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ and $\lambda_i > 0$ so that $T(\boldsymbol{e}_i) = \lambda_i \boldsymbol{v}_i$. Consider a covering $\boldsymbol{y}_1 + B_2^n, \dots, \boldsymbol{y}_N + B_2^n$ of $\mathcal{E}$ with $N := N(\mathcal{E}, B_2^n)$ many translates. Consider the inverse map T^{-1} with $T^{-1}(\boldsymbol{e}_i) = \frac{1}{\lambda_i}\boldsymbol{v}_i$. Then we can cover $T^{-1}(\mathcal{E}) = B_2^n$ with translates $T^{-1}(\boldsymbol{y}_i + B_2^n) = T^{-1}(\boldsymbol{y}_i) + \mathcal{E}^\circ$. $\square$



We will now show a first relation between covering numbers and volume — the big caveat is that the 2nd body T needs to be symmetric:

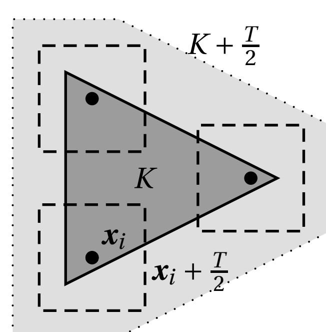
Lemma 4.3. *If K is convex and T is convex and symmetric, then*

$$\frac{Vol_n(K)}{Vol_n(T)} \leq N(K,T) \leq 2^n \frac{Vol_n(K + \frac{T}{2})}{Vol_n(T)}$$

The first inequality also holds without the symmetry assumption.

Proof. The first bound is trivial and clearly holds even if T is not symmetric. For the second inequality, select a *maximum* number of points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in K$ so that the translates $\boldsymbol{x}_i + \frac{T}{2}$ are disjoint. Then $K \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + T)$ since otherwise one could have added one more translate. Next, observe that $\boldsymbol{x}_i + \frac{T}{2} \subseteq K + \frac{T}{2}$ and hence

$$N(K,T) \leq N \overset{\text{disjointness}}{\leq} \frac{\mathrm{Vol}_n(K + \frac{T}{2})}{\mathrm{Vol}_n(\frac{T}{2})} = 2^n \frac{\mathrm{Vol}_n(K + \frac{T}{2})}{\mathrm{Vol}_n(T)}$$



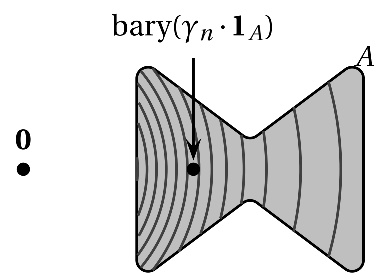
$\square$

4.2 The Milman-Pajor Theorem

In this section, we will finally prove that indeed $N(A,B) \approx \frac{\mathrm{Vol}_n(A-B)}{\mathrm{Vol}_n(B)}$ for all convex bodies. We begin with an estimate dealing with averages of log-concave functions. Let $A \subseteq \mathbb{R}^n$ be a measurable set (not necessarily convex), then the *bary center* is $\mathrm{bary}(A) = \mathbb{E}_{\boldsymbol{x}\sim A}[\boldsymbol{x}] = \frac{1}{\mathrm{Vol}_n(A)} \int_{\mathbb{R}^n} \boldsymbol{x} \cdot \mathbf{1}_A(\boldsymbol{x}) d\boldsymbol{x}$. Similarly if $F : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is just a non-negative function we can define a barycenter as

$$\mathrm{bary}(F) := \frac{1}{\int_{\mathbb{R}^n} F(\boldsymbol{x}) d\boldsymbol{x}} \int_{\mathbb{R}^n} \boldsymbol{x} \cdot F(\boldsymbol{x}) d\boldsymbol{x}$$

Note that in this notation $\mathrm{bary}(A) = \mathrm{bary}(\mathbf{1}_A)$ as one would expect. For example if the log-concave function is the density of the Gaussian restricted to A then the picture might look as follows:



Now suppose one is interested in the average value of a log concave function F over some set A. By concavity one would hope that this average might be bounded by the function value at the barycenter of F on A. And indeed, this is true.

We will translate the set and the function so that the barycenter is the origin (which simplifies the exposition).

Lemma 4.4. *Let $A \subseteq \mathbb{R}^n$ be a measurable set and let $F : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be log concave on A. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim A}[\boldsymbol{x} \cdot F(\boldsymbol{x})] = \mathbf{0} \quad \Rightarrow \quad \mathop{\mathbb{E}}_{\boldsymbol{x}\sim A}[F(\boldsymbol{x})] \leq F(\mathbf{0})$$

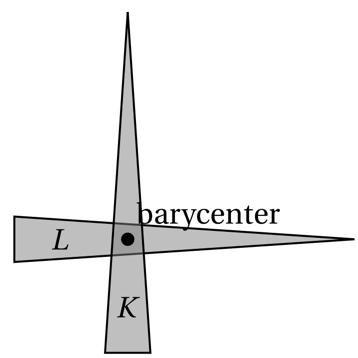
Proof 1. Let $M := \mathbb{E}_{\boldsymbol{x}\sim A}[F(\boldsymbol{x})]$ be the quantity that we try to upperbound. Moreover, let μ be the distribution with density $\frac{\mathbf{1}_A(\boldsymbol{x})}{\mathrm{Vol}_n(A)} \cdot \frac{F(\boldsymbol{x})}{M}$. Note that indeed this is a density as $\int_{\mathbb{R}^n} \mu(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{\mathrm{Vol}_n(A)\cdot M} \int_A F(\boldsymbol{x}) d\boldsymbol{x} = 1$. Also $\mathbb{E}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}] = \mathbb{E}_{\boldsymbol{x}\sim A}[\boldsymbol{x} \cdot \frac{F(\boldsymbol{x})}{M}] = \mathbf{0}$

by assumption. Then

$$\begin{aligned}
\ln(F(\mathbf{0})) &= \ln\Big(F\Big(\overbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}]}^{=\mathbf{0}}\Big)\Big) \\
&\overset{\text{Jensen on } \mu \text{ with } \ln(F) \text{ concave}}{\geq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[\ln(F(\boldsymbol{x}))] \\
&\overset{\text{Def } \mu}{=} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim A}\Big[\frac{F(\boldsymbol{x})}{M}\ln(F(\boldsymbol{x}))\Big] \\
&= \frac{1}{M}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim A}[F(\boldsymbol{x})\cdot\ln(F(\boldsymbol{x}))] \\
&\overset{\text{Jensen on } A \text{ with } t\mapsto t\ln(t) \text{ convex}}{\geq} \frac{1}{M}\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim A}[F(\boldsymbol{x})]}_{=M}\cdot\ln\Big(\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim A}[F(\boldsymbol{x})]}_{=M}\Big) = \ln(M)
\end{aligned}$$

Rearranging gives the claimed bound of $F(\mathbf{0}) \geq M = \mathbb{E}_{\boldsymbol{x}\sim A}[F(\boldsymbol{x})]$. $\square$

For two convex sets K, L it can happen that the intersection $K \cap L$ is very small. In that case one can prove that $K - L$ has to be quite large assuming they share the barycenter.



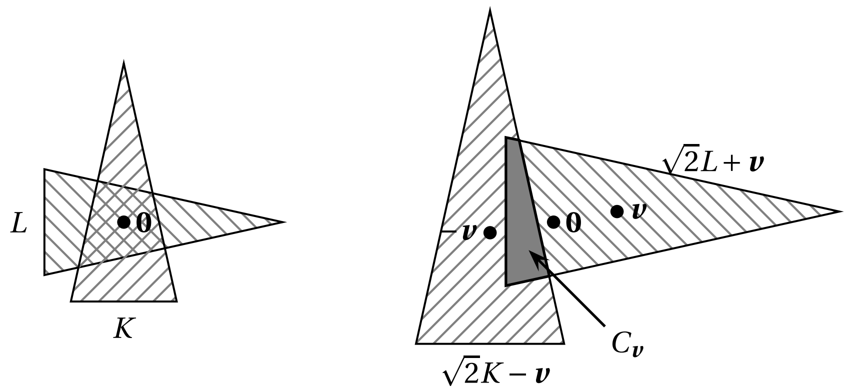
Theorem 4.5 (Milman-Pajor)**.** *Let $K, L \subseteq \mathbb{R}^n$ be convex bodies with the same barycenter. Then*

$$Vol_n(K)\cdot Vol_n(L) \leq Vol_n(K-L)\cdot Vol_n(K\cap L)$$

Proof. After translating both sets K and L by the same vector we may assume that $\text{bary}(K) = \mathbf{0} = \text{bary}(L)$. For a vector $\boldsymbol{v} \in \mathbb{R}^n$ we define

$$C_{\boldsymbol{v}} := (\sqrt{2}K - \boldsymbol{v}) \cap (\sqrt{2}L + \boldsymbol{v}) = \Big\{\boldsymbol{u}\in\mathbb{R}^n \mid \boldsymbol{u}+\boldsymbol{v}\in\sqrt{2}K \text{ and } \boldsymbol{u}-\boldsymbol{v}\in\sqrt{2}L\Big\}$$

as the shifted and scaled intersection.



Note that there is a linear bijective map between pairs $(\boldsymbol{x}, \boldsymbol{y})$ and $(\boldsymbol{u}, \boldsymbol{v})$ which is of the form

$$\boldsymbol{u} = \frac{\boldsymbol{x}+\boldsymbol{y}}{\sqrt{2}}, \boldsymbol{v} = \frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}} \quad \Leftrightarrow \quad \boldsymbol{x} = \frac{\boldsymbol{u}+\boldsymbol{v}}{\sqrt{2}}, \boldsymbol{y} = \frac{\boldsymbol{u}-\boldsymbol{v}}{\sqrt{2}}$$

The determinant of this map is 1, which leads to a simple change of variables in an integration while preserving the value of the integral.

We use this observation to rewrite the integral over a function $f : \mathbb{R}^n \to \mathbb{R}$ as

$$\begin{aligned} I(f) &:= \int_{K\times L} f\Big(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\Big) d\boldsymbol{x} d\boldsymbol{y} \\ &\overset{\boldsymbol{u}:=\frac{\boldsymbol{x}+\boldsymbol{y}}{\sqrt{2}}, \boldsymbol{v}:=\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}}{=} \int_{\mathbb{R}^n} f(\boldsymbol{v}) \cdot \mathrm{Vol}_n\Big(\Big\{\boldsymbol{u} \in \mathbb{R}^n \mid \frac{\boldsymbol{u}+\boldsymbol{v}}{\sqrt{2}} \in K \text{ and } \frac{\boldsymbol{u}-\boldsymbol{v}}{\sqrt{2}} \in L\Big\}\Big) d\boldsymbol{v} \\ &= \int_{\mathbb{R}^n} f(\boldsymbol{v}) \cdot \mathrm{Vol}_n(C_{\boldsymbol{v}}) d\boldsymbol{v} \quad (*) \end{aligned}$$

where in the 2nd step we made a simple change of variables. We choose set $Q := \frac{K-L}{\sqrt{2}}$ as the scaled Minkowski difference of both sets. The crucial part of the proof is to understand the function $F(\boldsymbol{v}) := \mathrm{Vol}_n(C_{\boldsymbol{v}})$. Next, we prove that the barycenter of the function F restricted to Q lies at the origin. For this same, we apply $(*)$ with the function $f(\boldsymbol{v}) := \boldsymbol{v} \cdot \mathbf{1}_Q(\boldsymbol{v})$ (note that the formula in $(*)$ also holds for vector-valued functions):

$$\begin{aligned} \int_{\mathbb{R}^n} \boldsymbol{v} \cdot \mathbf{1}_Q(\boldsymbol{v}) \cdot \mathrm{Vol}_n(C_{\boldsymbol{v}}) d\boldsymbol{v} &\overset{\text{apply } (*)}{=} \int_{K\times L} \Big(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\Big) \cdot \mathbf{1}_Q\Big(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\Big) d\boldsymbol{x} d\boldsymbol{y} \\ &= \frac{1}{\sqrt{2}} \int_K \boldsymbol{x} \cdot \Big(\int_L \mathbf{1}_Q\Big(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\Big) d\boldsymbol{y}\Big) d\boldsymbol{x} - \frac{1}{\sqrt{2}} \int_L \boldsymbol{y} \cdot \Big(\int_K \mathbf{1}_Q\Big(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\Big) d\boldsymbol{x}\Big) d\boldsymbol{y} \\ &= \underbrace{\frac{1}{\sqrt{2}} \int_K \boldsymbol{x} \cdot \underbrace{\mathrm{Vol}_n\Big(\Big\{\boldsymbol{y} \in L : \frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}} \in \frac{K-L}{\sqrt{2}}\Big\}\Big)}_{=\mathrm{Vol}_n(L)} d\boldsymbol{x}}_{=\mathbf{0}} - \underbrace{\frac{1}{\sqrt{2}} \int_L \boldsymbol{y} \cdot \underbrace{\mathrm{Vol}_n\Big(\Big\{\boldsymbol{x} \in K : \frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}} \in \frac{K-L}{\sqrt{2}}\Big\}\Big)}_{=\mathrm{Vol}_n(K)} d\boldsymbol{y}}_{=\mathbf{0}} \end{aligned}$$

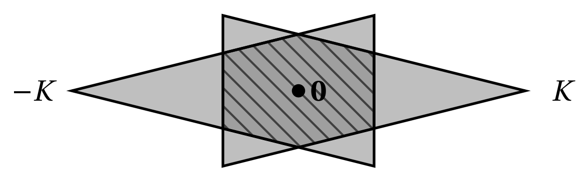
We conclude that indeed $\mathrm{bary}(\mathbf{1}_Q \cdot F) = \mathbf{0}$. Now we apply $(*)$ again — this time for the function $f(\boldsymbol{v}) := \mathbf{1}_Q(\boldsymbol{v})$. Then

$$\begin{aligned}
\mathrm{Vol}_n(K)\cdot\mathrm{Vol}_n(L) \quad &\leq \quad \int_{K\times L} \overbrace{\mathbf{1}_Q\Big(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\Big)}^{=1 \text{ if } \boldsymbol{x}\in K \text{ and } \boldsymbol{y}\in L} \; d\boldsymbol{x}d\boldsymbol{y} \\
&\stackrel{\text{apply } (*) \text{ with } f:=1}{=} \quad \int_{\mathbb{R}^n} \mathbf{1}_Q(\boldsymbol{v})\cdot \mathrm{Vol}_n(C_{\boldsymbol{v}})d\boldsymbol{v} = \mathrm{Vol}_n(Q)\cdot \mathop{\mathbb{E}}_{\boldsymbol{v}\sim Q}\big[\mathrm{Vol}_n(C_{\boldsymbol{v}})\big] \\
&\stackrel{\text{Lem 4.4}}{\leq} \quad \mathrm{Vol}_n(Q)\cdot\mathrm{Vol}_n(C_{\mathbf{0}}) = \mathrm{Vol}_n\Big(\frac{K-L}{\sqrt{2}}\Big)\cdot\mathrm{Vol}_n\Big(\sqrt{2}\cdot(K\cap L)\Big) \\
&= \quad \mathrm{Vol}_n(K-L)\cdot\mathrm{Vol}_n(K\cap L).
\end{aligned}$$

Here we use that the function $F(\boldsymbol{v}) := \mathrm{Vol}_n(C_{\boldsymbol{v}})$ is log concave by Lemma 1.34. □

A useful consequence is the following:

Corollary 4.6. *Let $K \subseteq \mathbb{R}^n$ be a convex body with bary$(K) = \mathbf{0}$. Then Vol$_n(K\cap(-K)) \geq 2^{-n}Vol_n(K)$.*



Proof. Applying the last Lemma with K and $L := -K$ gives

$$\mathrm{Vol}_n(K)\cdot\mathrm{Vol}_n(-K) \stackrel{(**)}{\leq} \underbrace{\mathrm{Vol}_n(K-(-K))}_{=\mathrm{Vol}_n(2K)}\cdot\mathrm{Vol}_n(K\cap(-K)) = 2^n\cdot\mathrm{Vol}_n(K)\cdot\mathrm{Vol}_n(K\cap(-K))$$

where we can apply Theorem 4.5 in $(**)$ as $\mathrm{bary}(K) = \mathrm{bary}(-K)$. □

In particular, the corollary shows that any convex set K contains a *symmetric* convex body as a subset that has at least a $2^{-\Theta(n)}$ fraction of the volume.

With Theorem 4.5 we can finally prove one of the main results of this Chapter:

Theorem 4.7. *Let $K, L \subseteq \mathbb{R}^n$ be convex bodies. Then*

$$4^{-n}\frac{Vol_n(K-L)}{Vol_n(L)} \leq N(K,L) \leq 4^n\frac{Vol_n(K-L)}{Vol_n(L)}$$

Proof. Translate L so that the barycenter of L is at the origin. The lower bound on the covering number

$$\begin{aligned} N(K,L) &\overset{\text{Rogers-Shephard}}{\geq} \overbrace{N(K-L,L-L)}^{\leq N(K,L)} \cdot \overbrace{4^{-n} \cdot \frac{\mathrm{Vol}_n(L-L)}{\mathrm{Vol}_n(L)}}^{\leq 1} \\ &\overset{\text{Lem 4.3}}{\geq} 4^{-n} \frac{\mathrm{Vol}_n(K-L)}{\mathrm{Vol}_n(L-L)} \cdot \frac{\mathrm{Vol}_n(L-L)}{\mathrm{Vol}_n(L)} = 4^{-n} \frac{\mathrm{Vol}_n(K-L)}{\mathrm{Vol}_n(L)} \end{aligned}$$

Here for the 1st inequality we use the Rogers-Shephard Inequality (Theorem 1.47) as well as the following observation: If $K \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + L)$ is an covering of K, then also $K - L \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + (L-L))$.

For the upper bound we simply use the best covering using the "symmetrizer" $S := L \cap (-L)$. Then

$$N(K,L) \leq N(K,S) \overset{\text{Lem 4.3}}{\leq} \frac{\mathrm{Vol}_n(K + \frac{1}{2}S)}{\mathrm{Vol}_n(\frac{1}{2}S)} \overset{(*)}{\leq} 2^n \cdot 2^n \frac{\mathrm{Vol}_n(K-L)}{\mathrm{Vol}_n(L)}$$

where we use in $(*)$ that $(-\frac{1}{2})S \subseteq L$ and $\mathrm{Vol}_n(S) \geq 2^{-n}\mathrm{Vol}_n(L)$. $\square$

We also provide a useful estimate for the covering numbers between a convex body K and its "symmetrizers" $K \cap (-K)$ and $K - K$.

Lemma 4.8. *Let $K \subseteq \mathbb{R}^n$ be a convex body with bary$(K) = \mathbf{0}$. Then $N(K-K, K \cap (-K)) \leq 2^{5n}$ and hence $N(K-K,K) \leq 2^{5n}$ and $N(K, K \cap (-K)) \leq 2^{5n}$.*

Proof. It suffices to estimate that

$$\begin{aligned} N(K-K, K\cap(-K)) &\overset{\text{Lem 4.3}}{\leq} 2^n \frac{\mathrm{Vol}_n\big((K-K) + \frac{1}{2}(K \cap (-K))\big)}{\mathrm{Vol}_n(K \cap (-K))} \\ &\leq 2^n 2^n \frac{\mathrm{Vol}_n(K-K)}{\mathrm{Vol}_n(K\cap(-K))} \overset{\text{Thm. 1.47, Cor. 4.6}}{\leq} 2^n 2^n \frac{2^{2n}\mathrm{Vol}_n(K)}{2^{-n}\mathrm{Vol}_n(K)} = 2^{5n} \end{aligned}$$

$\square$

4.3 The Primal and Dual Sudakov Inequality

Recall that the *mean width* of a convex body K is $w(K) := \mathbb{E}_{\boldsymbol{a} \sim S^{n-1}}[w_K(\boldsymbol{a})]$, where the width of K in direction $\boldsymbol{a} \in S^{n-1}$ is denoted as $w_K(\boldsymbol{a}) := \max\{|\langle \boldsymbol{a}, \boldsymbol{x}\rangle - \langle \boldsymbol{a}, \boldsymbol{y}\rangle| : \boldsymbol{x}, \boldsymbol{y} \in K\}$. The Sudakov Inequality relates the covering number to the mean width of the body. To be more precise, we will see the "primal" Sudakov inequality which upperbounds $N(K, B_2^n)$ and the Dual Sudakov Inequality which upper bounds $N(B_2^n, K)$. Our proof strategy is to first prove the dual Sudakov Inequality (which is actually due to Pajor and Tomczak) and then transfer the result to the primal setting. The statements assume that K is convex and symmetric, but at least for the Primal Sudakov Inequality this is really without loss of generality as $w(\text{conv}(K)) = w(K)$ and $w(K-K) = 2w(K)$.

We begin with a well-known estimate how the Gaussian measure of a symmetric body behaves under translation.

Lemma 4.9. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then for any vector $\boldsymbol{z} \in \mathbb{R}^n$ one has $\gamma_n(K+\boldsymbol{z}) \geq e^{-\|\boldsymbol{z}\|_2^2/2} \cdot \gamma_n(K)$.*

Proof. We simply write

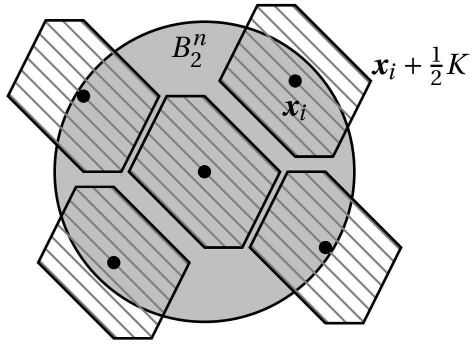
$$\begin{aligned}
\gamma_n(K+\boldsymbol{z}) &= \frac{1}{(2\pi)^{n/2}} \int_K \exp\Big(-\frac{1}{2}\|\boldsymbol{x}+\boldsymbol{z}\|_2^2\Big) d\boldsymbol{x} \\
&\stackrel{\text{symmetry}}{=} \frac{1}{(2\pi)^{n/2}} \int_K \mathop{\mathbb{E}}_{\sigma \in \{-1,1\}} \Big[\exp\Big(-\frac{1}{2}\|\sigma\boldsymbol{x}+\boldsymbol{z}\|_2^2\Big)\Big] d\boldsymbol{x} \\
&\stackrel{\text{Jensen}}{\geq} \frac{1}{(2\pi)^{n/2}} \int_K \exp\Big(-\frac{1}{2} \underbrace{\mathop{\mathbb{E}}_{\sigma \in \{-1,1\}} \big[\|\sigma\boldsymbol{x}+\boldsymbol{z}\|_2^2\big]}_{=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{z}\|_2^2}\Big) d\boldsymbol{x} \\
&= \exp\Big(-\frac{\|\boldsymbol{z}\|_2^2}{2}\Big) \cdot \gamma_n(K)
\end{aligned}$$

using Jensen's inequality with the fact that $f(\boldsymbol{y}) = \exp(-\frac{1}{2}\boldsymbol{y})$ is convex. □

Theorem 4.10 (Dual Sudakov Inequality - Pajor-Tomczak)**.** *Let $K \subseteq \mathbb{R}^n$ be centrally symmetric convex body. Then for any $t > 0$ one has*

$$N(B_2^n, tK) \leq \exp\Big(O(n) \cdot \Big(\frac{w(K^\circ)}{t}\Big)^2\Big)$$

Proof. We can scale K and t simultaneously until $t = 1$. Pick a maximal set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in B_2^n$ so that the translates $\boldsymbol{x}_i + \frac{1}{2}K$ are disjoint. Then by maximality, this indeed provides a covering of the ball in the sense that $B_2^n \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + K)$. In other words, $N(B_2^n, K) \leq N$.



It remains to bound the number N. For a parameter $\lambda \geq 0$ that we determine later, the Gaussian measure of these translates satisfies

$$
\begin{aligned}
1 &\overset{\text{measure}}{\geq} \gamma_n\Big(\bigcup_{i=1}^{N} \lambda \cdot \Big(\boldsymbol{x}_i + \frac{1}{2}K\Big)\Big) \overset{\text{disjointness}}{\geq} \sum_{i=1}^{N} \gamma_n\Big(\lambda \boldsymbol{x}_i + \frac{\lambda}{2}K\Big) \\
&\overset{\text{Lem 4.9}}{\geq} \sum_{i=1}^{N} e^{-\lambda^2 \|\boldsymbol{x}_i\|_2^2/2} \cdot \gamma_n\Big(\frac{\lambda}{2}K\Big) \overset{\|\boldsymbol{x}_i\|_2 \leq 1}{\geq} N \cdot e^{-\lambda^2/2} \cdot \gamma_n\Big(\frac{\lambda}{2}K\Big) \quad (*)
\end{aligned}
$$

This provides us with an upper bound of $N \leq \frac{\exp(\lambda^2/2)}{\gamma_n(\frac{\lambda}{2}K)}$ on the number of translates.

Next, we need to relate the Gaussian measure of K to $w(K^\circ)$. We write

$$
\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_K] = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_2] \cdot \mathop{\mathbb{E}}_{\boldsymbol{\theta}\sim S^{n-1}}[\underbrace{\|\boldsymbol{\theta}\|_K}_{=h_{K^\circ}(\boldsymbol{\theta})}] \leq \frac{\sqrt{n}}{2} \cdot w(K^\circ), \quad (**)
$$

see Lemma 1.1 and Lemma 1.4 (note that this is essentially an equality). Then applying Markov's inequality to $(**)$ gives $\Pr_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_K \geq \frac{\lambda}{2}] \leq \frac{\mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_K]}{\lambda/2} \leq \frac{\sqrt{n}}{\lambda} w(K^\circ)$ and so for a choice of $\lambda := 2\sqrt{n} \cdot w(K^\circ)$, we get $\gamma_n(\frac{\lambda}{2}K) \geq \frac{1}{2}$. Then for this choice of λ we get

$$
N(B_2^n, K) \leq N \leq \frac{\exp(\lambda^2/2)}{\gamma_n(\frac{\lambda}{2}K)} \leq \exp(\Theta(n) \cdot w(K^\circ)^2)
$$

as claimed. $\square$

We note that the translates choosen in the proof have their centers in B_2^n, hence we have actually proven the stronger claim of $\bar{N}(B_2^n, tK) \leq \exp\big(O(n) \cdot \big(\frac{w(K^\circ)}{t}\big)^2\big)$.

In order to derive the "primal" Sudakov Inequality we need to have a relation between covering numbers $N(K, B_2^n)$ and $N(B_2^n, K^\circ)$. Note that we anyway have $N(2tB_2^n, tB_2^n) \leq 2^{O(n)}$ and hence $N(K, tB_2^n) \leq 2^{O(n)} N(K, 2tB_2^n)$. However the following lemma gives us a finer control where we might not have to pay that exponential factor.

Lemma 4.11. *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex body. Then*

$$N(K, tB_2^n) \leq N(K, 2tB_2^n) \cdot N\Big(B_2^n, \frac{t}{8}K^\circ\Big)$$

Proof. The claim is invariant under scaling, so it suffices to prove the claim for $t = 1$, i.e. we show $N(K, B_2^n) \leq N(K, 2B_2^n) \cdot N(B_2^n, \frac{1}{8}K^\circ)$. First observe that $2K \cap \frac{1}{2}K^\circ \subseteq B_2^n$ as for $\boldsymbol{x} \in (2K \cap \frac{1}{2}K^\circ)$ one has $\|\boldsymbol{x}\|_2^2 = \langle \boldsymbol{x}, \boldsymbol{x}\rangle \leq \|\boldsymbol{x}\|_K \cdot \|\boldsymbol{x}\|_{K^\circ} \leq 2 \cdot \frac{1}{2} = 1$. Then

$$N(K, B_2^n) \overset{2K\cap\frac{1}{2}K^\circ\subseteq B_2^n}{\leq} N\Big(K, \frac{1}{2}K^\circ \cap 2K\Big) \overset{(*)}{\leq} N\Big(K, \frac{1}{4}K^\circ\Big) \leq N(K, 2B_2^n) \cdot \underbrace{N\Big(2B_2^n, \frac{1}{4}K^\circ\Big)}_{=N(B_2^n, \frac{1}{8}K^\circ)}$$

Here we use in $(*)$ that by Lemma 4.1.(5), for two symmetric convex sets $A, B \subseteq \mathbb{R}^n$ one has $N(A, 2(A \cap B)) \leq N(A, B)$. □

Now we can prove Sudakov's inequality. Note that this inequality works also for non-symmetric bodies:

Theorem 4.12 (Sudakov Inequality)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then for $t > 0$ one has*

$$N(K, tB_2^n) \leq \exp\Big(\Theta(n) \cdot \Big(\frac{w(K)}{t}\Big)^2\Big)$$

Proof. Translate K so that $\mathbf{0} \in K$. Note that $K \subseteq K - K$ and $K - K$ is a centrally symmetric convex body with $w(K - K) = 2w(K)$. Hence it suffices to prove the inequality for the symmetric case. So suppose that K is symmetric. First we obtain that

$$N(K, tB_2^n) \overset{\text{Lem 4.11}}{\leq} N(K, 2tB_2^n) \cdot N\Big(B_2^n, \frac{t}{8}K^\circ\Big) \overset{\text{Thm 4.10}}{\leq} N(K, 2tB_2^n) \cdot e^{Cnw(K)^2/t^2} \quad (*)$$

by applying the Theorem of Pajor-Tomjak to the body K° and using that $(K^\circ)^\circ = K$. Let $A := \sup_{t>0}\{t^2 \ln N(K, tB_2^n)\}$. Note that this is the minimal quantity so that $N(K, tB_2^n) \leq \exp(A/t^2)$ for all $t > 0$[1]. For the sake of simplicity suppose this sup is attained. Then for this t one has

$$\begin{aligned} A &= t^2 \ln N(K, tB_2^n) \overset{(*)}{\leq} t^2 \cdot \Big(\ln N(K, 2tB_2^n) + \frac{Cn \cdot w(K)^2}{t^2}\Big) \\ &= \frac{1}{4}\underbrace{(2t)^2 \ln N(K, 2t \cdot B_2^n)}_{\leq A} + Cn \cdot w(K)^2 \leq \frac{A}{4} + Cn \cdot w(K)^2 \end{aligned}$$

[1] To see that this supremum is finite consider the following: Let r be the radius of K. Then for $t \geq r$ we have $N(K, tB_2^n) = 1 \leq \exp(A/t^2)$ no matter what $A \geq 0$ is. On the other hand, for $t < r$ we can use $N(K, B_2^n) \leq N(K, rB_2^n) \cdot N(rB_2^n, tB_2^n) \leq \exp(n\frac{Cr}{t})$.

which can be rearranged to $A \le \frac{4}{3}Cn \cdot w(K)^2$. This shows the claim. $\square$

We should point out that Sudakov's Inequality is only interesting for some specific regimes. For example by volume arguments we know that for $s \geq 1$ one has $N(sB_2^n, B_2^n) \leq O(s)^n$. On the other hand, Sudakov's Inequality provides us with a rather disappointing bound of $N(sB_2^n, B_2^n) \leq \exp(O(ns^2))$.

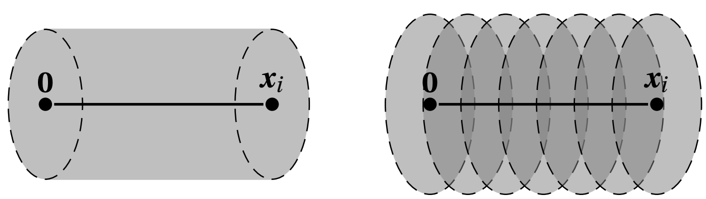
4.4 Additional estimates on covering numbers

We want to discuss several other estimates on volumes and covering numbers that will in particular useful in later chapters. First we show that covering conv$(K \cup L)$ takes basically $N(L,K)$ many translates of K, times a linear factor.

Lemma 4.13. *Let $K, L \subseteq \mathbb{R}^n$ be convex bodies where K is symmetric and $L \subseteq \beta \cdot K$ for some $\beta \geq 1$. Then $N(conv(K \cup L), (1+\frac{1}{n}) \cdot K) \leq 2\beta n \cdot N(L,K)$.*

Proof. Suppose that $L \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + K)$ is the minimal covering of L with translates of K. We may assume that $L \cap (\boldsymbol{x}_i + K) \neq \emptyset$ and hence $\boldsymbol{x}_i \in L + K \subseteq 2\beta K$. Then

$$\mathrm{conv}(L \cup K) \subseteq \bigcup_{\substack{i \in [N],\\ 0 \le \lambda \le 1}} (\lambda(\boldsymbol{x}_i + K) + (1-\lambda)K) = \bigcup_{\substack{i \in [N],\\ 0 \le \lambda \le 1}} (\lambda \boldsymbol{x}_i + K) \subseteq \bigcup_{\substack{i \in [N],\\ 0 \le \lambda \le 1,\\ \lambda \in \frac{\mathbb{N}}{2\beta n}}} \left(\lambda \boldsymbol{x}_i + \left(1 + \frac{1}{n}\right)K\right)$$



$$\bigcup_{0 \le \lambda \le 1} (\lambda \boldsymbol{x}_i + K) \qquad \bigcup_{0 \le \lambda \le 1, \lambda \in \frac{\mathbb{N}}{2\beta n}} \left(\lambda \boldsymbol{x}_i + \left(1 + \frac{1}{n}\right)K\right)$$

Here we use for the last inclusion that for $|\lambda - \lambda'| \leq \frac{1}{2\beta n}$ one has $\|\lambda \boldsymbol{x}_i - \lambda' \boldsymbol{x}_i\|_K \leq \frac{1}{n}$. $\square$

We can easily turn the last estimate into a volume bound. Here it also becomes more clear that we were satisfied with the $(1+\frac{1}{n})$-blowup of K as this means only a constant factor blowup of the volume.

Lemma 4.14. *Let $L, K \subseteq \mathbb{R}^n$ be a convex bodies where K is also symmetric. Suppose that $L \subseteq \beta K$ for some $\beta \geq 1$. Then*

$$Vol_n(conv(K \cup L)) \leq 6\beta n \cdot N(L,K) \cdot Vol_n(K).$$

Proof. Using Lemma 4.13 we conclude that

$$\mathrm{Vol}_n(\mathrm{conv}(K\cup L)) \leq \underbrace{N\big(\mathrm{conv}(K\cup L),(1+\tfrac{1}{n})K\big)}_{\leq 2\beta n\cdot N(L,K)}\cdot\underbrace{\mathrm{Vol}_n\big((1+\tfrac{1}{n})K\big)}_{\leq 3\cdot\mathrm{Vol}_n(K)}$$

□

We also prove two Lemmas that we need in particular in Chapter 8.

Lemma 4.15. *For symmetric convex bodies $K,P\subseteq\mathbb{R}^n$ one has*

$$Vol_n(K+P)\leq Vol_n((rB_2^n\cap K)+P)\cdot N(K,rB_2^n)$$

Proof. After scaling we may assume $r=1$. Consider the set $T(\boldsymbol{x}):=((\boldsymbol{x}+B_2^n)\cap K)+P$ and notice that by the symmetry and convexity of K,P,B_2^n we have $\frac{1}{2}T(\boldsymbol{x})+\frac{1}{2}T(-\boldsymbol{x})\subseteq T(\mathbf{0})$. Then by the Brunn-Minkowski Inequality, one has $\mathrm{Vol}_n(T(\boldsymbol{x}))\leq\mathrm{Vol}_n(T(\mathbf{0}))$. Set $N:=N(K,B_2^n)$ and consider the covering $K\subseteq\bigcup_{i=1}^N((\boldsymbol{x}_i+B_2^n)\cap K)$. Then clearly

$$K+P\subseteq\bigcup_{i=1}^N((\boldsymbol{x}_i+B_2^n)\cap K)+P$$

hence $\mathrm{Vol}_n(K+P)\leq\sum_{i=1}^N\mathrm{Vol}_n(T(\boldsymbol{x}_i))\leq N\cdot T(\mathbf{0})$ which gives the claim. □

We will also need an extension of Lemma 4.14 for Chapter 8:

Lemma 4.16. *Let $K,P\subseteq\mathbb{R}^n$ be symmetric convex bodies with $rB_2^n\subseteq\beta K$ for some $\beta\geq 1$ and $r>0$, then*

$$Vol_n\big(conv(K\cup rB_2^n)+P\big)\leq 6\beta n\cdot N(rB_2^n,K)\cdot Vol_n(K+P).$$

Proof. After scaling, $r=1$. We have seen in Lemma 4.13 that there are is a covering with $N\leq 2\beta n\cdot N(B_2^n,K)$ points so that $\mathrm{conv}(K\cup B_2^n)\subseteq\bigcup_{i=1}^N(\boldsymbol{x}_i+(1+\frac{1}{n})K)$. Taking the volume of the Minkowski sum with P on both sides gives

$$\mathrm{Vol}_n(\mathrm{conv}(K\cup B_2^n)+P)\leq\sum_{i=1}^N\mathrm{Vol}_n\Big(\Big(\boldsymbol{x}_i+\Big(1+\frac{1}{n}\Big)K\Big)+P\Big)\leq\underbrace{\Big(1+\frac{1}{n}\Big)^n}_{\leq 3}\cdot N\cdot\mathrm{Vol}_n(K+P)$$

□

Finally we state a rather deep result without proof. It says that the covering numbers $N(K,B_2^n)$ and $N(B_2^n,K^\circ)$ are approximately equal:

Theorem 4.17 (Duality of Covering Numbers - Artstein-Milman-Szarek)**.** *For any symmetric convex body $K\subseteq\mathbb{R}^n$ one has $N(B_2^n,128K^\circ)^{1/10}\leq N(K,B_2^n)\leq N\big(B_2^n,\frac{1}{128}K^\circ\big)^{10}$.*

In fact, in Chapter 7 we will give a full proof for a similar relation that holds even if B_2^n is replaced by an arbitrary symmetric convex set and its dual.

4.5 Exercises

Exercise 4.1.
Prove the non-trivial part of Lemma 4.1.(1): For any convex bodies $A, B \subseteq \mathbb{R}^n$ one has $\bar{N}(A, B - B) \leq N(A, B)$.

Exercise 4.2.
Prove Lemma 4.1.(2): For any convex body $A \subseteq \mathbb{R}^n$ and $r > 0$ one has $\bar{N}(A, rB_2^n) \leq N(A, rB_2^n)$.

Exercise 4.3.
Let $A, B \subseteq \mathbb{R}^n$ be convex bodies where at least B is symmetric. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in A$ be any maximal set of points so that the translates $\boldsymbol{x}_1 + B, \dots, \boldsymbol{x}_N + B$ are disjoint. Prove that $N(A, 2B) \leq N \leq N(A, B)$.

Exercise 4.4.
Prove that any symmetric convex bodies $A, B \subseteq \mathbb{R}^n$ satisfy $N(A, B) \leq (Cn)^n \cdot N(B^\circ, A^\circ)$ for a universal constant $C > 0$.

Exercise 4.5.
Let $K \subseteq \mathbb{R}^n$ be any convex body. Prove that $N(K - K, K) \leq C^n$ for some universal constant $C > 0$.

Exercise 4.6.
Let $t > 0$. What estimate does the Dual Sudakov Inequality provide for the quantity $N(B_2^n, tB_\infty^n)$?

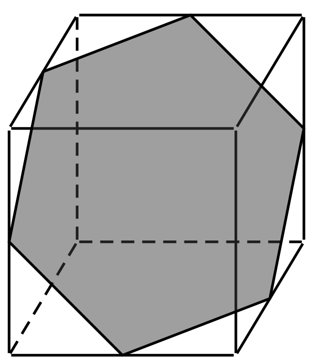
Chapter 5

Almost Euclidean Subspaces of finite dimensional normed spaces

In this chapter we discuss a remarkable Theorem of Dvoretzky [Dvo59, Dvo61] where the quantitative bound that we present is due to Milman [Mil71]:

> *For every symmetric convex body $K \subseteq \mathbb{R}^n$ there is a subspace $V \subseteq \mathbb{R}^n$ of dimension $\Omega(\frac{\varepsilon^2}{\log(1/\varepsilon)}\log(n))$ so that $\|\boldsymbol{x}\|_K = (1\pm\varepsilon)\cdot R\cdot\|\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in V$ where $R \in \mathbb{R}_{>0}$ is a suitable radius.*

For example the cube $[-1,1]^3$ intersected with a hyperplane already looks a little more spherical:



We introduce some notation that we use later in the chapter. The *M-value* of a symmetric convex body $K \subseteq \mathbb{R}^n$ is

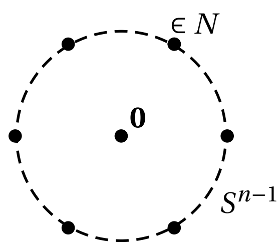
$$M(K) := \mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_K]$$

Note that the bigger K is, the smaller is $M(K)$. Also we should point out that this is not actually a new quantity as $2M(K) = w(K^\circ)$ by Theorem 1.4. But for the purpose of this chapter, it will be more natural to use $M(K)$ than $w(K^\circ)$.

5.1 Dvoretzky's Theorem

The overall strategy to prove Dvoretzky's Theorem will be to take a random subspace $V \subseteq \mathbb{R}^n$ of dimension $k = \Theta_\varepsilon(\log n)$ and then argue that $V \cap K$ is close to a ball in every direction. In general, we say that a a symmetric convex body $K \subseteq \mathbb{R}^n$ is $(1+\varepsilon)$-*spherical* if $\frac{1}{1+\varepsilon}\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq (1+\varepsilon)\|\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Moreover, we say that a symmetric convex body K is $(1+\varepsilon)$-*ellipsoidal* if there is an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ with $\frac{1}{1+\varepsilon}\|\boldsymbol{x}\|_{\mathcal{E}} \leq \|\boldsymbol{x}\|_K \leq (1+\varepsilon)\|\boldsymbol{x}\|_{\mathcal{E}}$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Then our goal can be rephrased to finding a subspace V so that $V \cap K$ is $(1+\varepsilon)$-spherical with respect to the ambient space V.

The first ingredient that we need for the proof of Dvoretzky's Theorem is the existence of small ε-nets. Here, we say that N is a ε-*net* for S^{n-1} if $N \subseteq S^{n-1}$ and for every point $\boldsymbol{x} \in S^{n-1}$ there exists a point $\boldsymbol{y} \in N$ so that $\|\boldsymbol{x} - \boldsymbol{y}\|_2 < \varepsilon$.



Lemma 5.1. *For any $0 < \varepsilon \leq 1$ there is a ε-net $N \subseteq S^{n-1}$ of size $|N| \leq (\frac{4}{\varepsilon})^n$.*

Proof. Pick any *maximal* set of points $N \subseteq S^{n-1}$ that have $\|\cdot\|_2$-distance at least ε to each other. Then N is a ε-net. Moreover the balls $\boldsymbol{x} + \frac{\varepsilon}{2}B_2^n$ are disjoint for $\boldsymbol{x} \in N$ and contained in $(1+\frac{\varepsilon}{2})B_2^n$. Hence

$$|N| \leq \frac{\mathrm{Vol}_n((1+\frac{\varepsilon}{2}) \cdot B_2^n)}{\mathrm{Vol}_n(\frac{\varepsilon}{2} \cdot B_2^n)} \leq \left(\frac{4}{\varepsilon}\right)^n$$

□

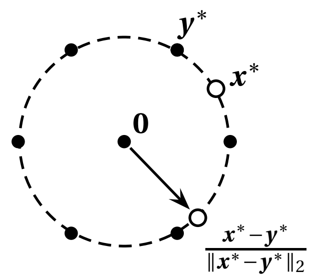
The next observation is that a body K is $(1+O(\varepsilon))$-spherical as soon as the condition $\frac{1}{1+\varepsilon}\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq (1+\varepsilon)\|\boldsymbol{x}\|_2$ is satisfied for all vectors in an ε-net.

Lemma 5.2. *Let $\|\cdot\|_K$ be a norm and let N be an ε-net for $\varepsilon \leq \frac{1}{4}$. Suppose that $(1-\varepsilon) \cdot R \leq \|\boldsymbol{y}\|_K \leq (1+\varepsilon) \cdot R$ for every $\boldsymbol{y} \in N$. Then $\|\boldsymbol{x}\|_K \in [1-3\varepsilon, 1+3\varepsilon] \cdot R$ for all $\boldsymbol{x} \in S^{n-1}$.*

Proof. Let $\boldsymbol{x}^* \in S^{n-1}$ be the point maximizing $\|\cdot\|_K$. Let $\boldsymbol{y}^* \in N$ be a point in the net with $\|\boldsymbol{x}^* - \boldsymbol{y}^*\|_2 \leq \varepsilon$. Then

$$\|\boldsymbol{x}^*\|_K \leq \|\boldsymbol{y}^*\|_K + \underbrace{\|\boldsymbol{x}^* - \boldsymbol{y}^*\|_K}_{\leq \varepsilon \|\boldsymbol{x}^*\|_K} \leq R(1+\varepsilon) + \varepsilon\|\boldsymbol{x}^*\|_K$$

where we use that $\|\boldsymbol{v}\|_K \leq \|\boldsymbol{v}\|_2 \cdot \|\boldsymbol{x}^*\|_K$ for any $\boldsymbol{v}$, by the choice of $\boldsymbol{x}^*$. Rearranging gives $\|\boldsymbol{x}^*\|_K \leq R \cdot \frac{1+\varepsilon}{1-\varepsilon} \leq R \cdot (1+3\varepsilon)$.



For the lower bound let $\boldsymbol{x}^{**} \in S^{n-1}$ be the vector *minimizing* $\|\cdot\|_K$ and let $\boldsymbol{y}^{**} \in N$ be a point with $\|\boldsymbol{x}^{**} - \boldsymbol{y}^{**}\|_2 \leq \varepsilon$. Then very similarly

$$\|\boldsymbol{x}^{**}\|_K \geq \|\boldsymbol{y}^{**}\|_K - \underbrace{\|\boldsymbol{x}^{**} - \boldsymbol{y}^{**}\|_K}_{\leq \varepsilon \|\boldsymbol{x}^*\|_K} \geq R \cdot (1-\varepsilon) - \varepsilon R(1+3\varepsilon) \geq R \cdot (1-3\varepsilon)$$

□

Next, we should discuss how one can actually generate a random subspace. Recall that a matrix $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ is called *orthogonal* if the column vectors are orthogonal unit vectors and the row vectors are orthogonal unit vectors. Moreover, recall that $O(n) := \{\boldsymbol{U} \in \mathbb{R}^{n\times n} \mid \boldsymbol{U} \text{ is orthogonal}\}$ is the set of orthogonal matrices, also called the *orthogonal group*. We write $\boldsymbol{U} \sim O(n)$ if we draw an orthogonal matrix uniform at random. Note that, if $\boldsymbol{x} \in S^{n-1}$ is any fixed unit vector and $\boldsymbol{U} \sim O(n)$, then $\boldsymbol{U}\boldsymbol{x} \sim S^{n-1}$ is uniformly distributed. More generally, if $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k \in \mathbb{R}^n$ are the first k standard basis vectors, then for $\boldsymbol{U} \sim O(n)$ we know that $\text{span}\{\boldsymbol{U}\boldsymbol{e}_1, \ldots, \boldsymbol{U}\boldsymbol{e}_k\}$ is a uniform random k-dimensional subspace. Recall that $M(K) = \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_K]$. We also want to remind the reader of a concentration result from Chapter 3, see Lemma 3.5. If $f : S^{n-1} \to \mathbb{R}$ is a Lipschitz function, then $\Pr_{\boldsymbol{x}\sim S^{n-1}}[|f(\boldsymbol{x}) - \mu| > t] \leq 64\exp(-nt^2/64)$ for any $t \geq 0$ where $\mu := \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[f(\boldsymbol{x})]$.

Theorem 5.3. *Let $K \subseteq \mathbb{R}^n$ be a convex symmetric body with $B_2^n \subseteq K$ and let $c > 0$ be a small enough constant and abbreviate $M := M(K)$. Let $V \subseteq \mathbb{R}^n$ be a uniform random subspace with* $\dim(V) = k$ *where $k \leq c\frac{\varepsilon^2}{\ln(\frac{1}{\varepsilon})} nM^2$. Then*

$$\Pr_V\left[(1-\varepsilon)\cdot M \cdot \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq (1+\varepsilon)\cdot M \cdot \|\boldsymbol{x}\|_2 \;\; \forall \boldsymbol{x} \in V\right] \geq \frac{1}{2}$$

Proof. Consider the function $f : S^{n-1} \to \mathbb{R}_{\geq 0}$ with $f(\boldsymbol{x}) := \|\boldsymbol{x}\|_K$. We can show the following:

Claim. One has $\Pr_{\boldsymbol{x}\sim S^{n-1}}[|f(\boldsymbol{x}) - M| > \frac{\varepsilon}{3}M] \leq 64\exp(-n\varepsilon^2 M^2/600)$.

Proof of claim. Recall that $\mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[f(\boldsymbol{x})] = M$. Next, we have $B_2^n \subseteq K$ and hence f is 1-Lipschitz as one can easily verify: $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq \|\boldsymbol{x} - \boldsymbol{y}\|_K \leq \|\boldsymbol{x} - \boldsymbol{y}\|_2$. By concentration (see Lemma 3.5)

$$\Pr_{\boldsymbol{x}\sim S^{n-1}}\Big[|f(\boldsymbol{x}) - M| > \frac{\varepsilon}{3}M\Big] \leq 64\exp\Big(-n\cdot\Big(\frac{\varepsilon M}{3}\Big)^2/64\Big) \quad \square$$

Now, fix a dimension k. Let N be an $\frac{\varepsilon}{3}$-net for $\mathbb{R}^k$ with $|N| \leq (12/\varepsilon)^k$ by Lemma 5.1. We write $N = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{|N|}\}$ where $\boldsymbol{y}_i \in \mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k\}$ and $\boldsymbol{e}_i \in \mathbb{R}^n$ is one of the standard basis vectors. Let $\boldsymbol{U} \in O(n)$ be a random orthogonal matrix. Then we can sample the random k-dimensional subspace as $V := \mathrm{span}\{\boldsymbol{U}\boldsymbol{e}_1, \ldots, \boldsymbol{U}\boldsymbol{e}_k\}$. Then for each $\boldsymbol{y}_i \in N$ we know that $\boldsymbol{U}\boldsymbol{y}_i$ is a uniform choice from S^{n-1} and hence using the union bound

$$\begin{aligned}
\Pr\Big[\underbrace{\Big(1-\frac{\varepsilon}{3}\Big)M \leq f(\boldsymbol{U}\boldsymbol{y}) \leq \Big(1+\frac{\varepsilon}{3}\Big)M \ \ \forall \boldsymbol{y} \in N}_{\text{event } (*)}\Big] &\overset{\text{Claim}}{\geq} 1 - |N|\cdot 64\exp\big(-n\cdot\varepsilon^2 M^2/600\big) \\
&= 1 - (12/\varepsilon)^k \cdot 64\exp\big(-n\cdot\varepsilon^2 M^2/600\big) \\
&= 1 - 64\exp\Big(k\ln\Big(\frac{12}{\varepsilon}\Big) - n\cdot\frac{\varepsilon^2 M^2}{600}\Big) \geq \frac{1}{2}
\end{aligned}$$

where the last inequality holds for $k \leq c\frac{\varepsilon^2}{\ln(\frac{1}{\varepsilon})}nM^2$ for a small enough constant $c > 0$. If event $(*)$ happens, then by Lemma 5.2 we have that $(1-\varepsilon)M \leq \|\boldsymbol{x}\|_K \leq (1+\varepsilon)M$ for all $\boldsymbol{x} \in V \cap S^{n-1}$, which then gives the result. $\square$

We should check what we can infer from the last estimate. Take a symmetric convex body K and assume that it is in John's position. Then $B_2^n \subseteq K \subseteq \sqrt{n}B_2^n$ and hence $M(K) \geq \frac{1}{\sqrt{n}}$. Then even if we are satisfied with a modest choice of $\varepsilon = \Theta(1)$, the condition from the last Lemma only works up to $k = \Theta(\varepsilon^2\ln(1/\varepsilon)) = \Theta(1)$ which is a rather vacuous statement. The weakpoint in that argument is that we used the simple estimate $M(K) \geq \frac{1}{\sqrt{n}}$. If that inequality was tight (say up to constant factors), then this would mean that for a body K in John's position one might have $\Pr_{\boldsymbol{x}\in S^{n-1}}[\rho_K(\boldsymbol{x}) \geq \Omega(\sqrt{n})] \geq \Omega(1)$ where $\rho_K(\boldsymbol{x})$ is again the radius of K in direction $\boldsymbol{x}$. Indeed we can prove that this pathological situation cannot occur:

Theorem 5.4. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body in John's position. Then one has $M(K) \geq \Omega\Big(\sqrt{\frac{\log(n)}{n}}\Big)$.*

Proof. We apply the *Dvoretzky-Rogers Theorem* (Lemma 2.9) to obtain an orthonormal basis $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ so that $\frac{1}{4} \leq \|\boldsymbol{x}_i\|_K \leq 1$. Then

$$\begin{aligned} M(K) &= \mathop{\mathbb{E}}_{\boldsymbol{a}\sim S^{n-1}}[\|\boldsymbol{a}\|_K] = \mathop{\mathbb{E}}_{\boldsymbol{a}\sim S^{n-1}}\Big[\Big\|\sum_{i=1}^n a_i\boldsymbol{x}_i\Big\|_K\Big] = \mathop{\mathbb{E}}_{\boldsymbol{a}\sim S^{n-1}}\Big[\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{\varepsilon}\sim\{-1,1\}^n}\Big[\Big\|\sum_{i=1}^n \varepsilon_i a_i\boldsymbol{x}_i\Big\|_K\Big]}_{\geq \|a_j\boldsymbol{x}_j\|_K\ \forall j}\Big] \\ &\stackrel{(***)}{\geq} \mathop{\mathbb{E}}_{\boldsymbol{a}\sim S^{n-1}}\Big[\max_{j=1,\dots,n}\big\|a_j\boldsymbol{x}_j\big\|_K\Big] \stackrel{\|\boldsymbol{x}_j\|_K\geq\Omega(1)}{\geq} \mathop{\mathbb{E}}_{\boldsymbol{a}\sim S^{n-1}}[\|\boldsymbol{a}\|_\infty] \\ &= \Theta\Big(\frac{1}{\sqrt{n}}\Big)\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{g}\sim\gamma_n}[\|\boldsymbol{g}\|_\infty]}_{=\Theta(\sqrt{\log(n)})} \geq \Omega\Big(\sqrt{\frac{\log(n)}{n}}\Big) \end{aligned}$$

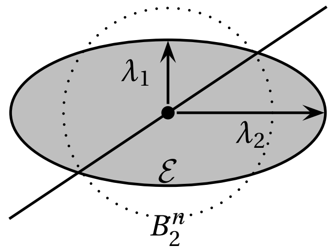
where we use that randomly flipping signs of $\boldsymbol{a}\sim S^{n-1}$ is not changing the distribution. We need to justify $(***)$. First note that $\mathbb{E}_{\sigma\in\{-1,1\}}[\|\boldsymbol{y}+\sigma\boldsymbol{x}\|_K] \geq \|\boldsymbol{x}\|_K$ for all $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$. To see this inequality write

$$\|\boldsymbol{x}\|_K = \Big\|\frac{1}{2}(\boldsymbol{y}+\boldsymbol{x}) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})\Big\|_K \leq \frac{1}{2}\|\boldsymbol{y}+\boldsymbol{x}\|_K + \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_K = \mathop{\mathbb{E}}_{\sigma\in\{-1,1\}}[\|\boldsymbol{y}+\sigma\boldsymbol{x}\|_K]$$

using the triangle inequality. To then finish $(***)$, fix an outcome for $\boldsymbol{a}\in S^{n-1}$ and the index $j\in[n]$ that maximizes $\|a_j\boldsymbol{x}_j\|_K$. Moreover fix the signs $\varepsilon_1,\dots,\varepsilon_{j-1},\varepsilon_{j+1},\dots,\varepsilon_n$ and set $\boldsymbol{y} := \sum_{i\neq j} a_i\boldsymbol{x}_i$ and $\boldsymbol{x} := a_j\boldsymbol{x}_j$. Finally we used that the maximum of a standard Gaussian in n dimensions is $\Theta(\sqrt{\log n})$. We take this as given for now and will prove it in a later Chapter. $\square$

In particular this implies that if K is a symmetric convex body in John's position, then a random subspace of dimension $k = \Theta(\frac{\varepsilon^2}{\ln(1/\varepsilon)}\log(n))$ is $(1+\varepsilon)$-spherical with high probability. Of course we also want to obtain almost-spherical subspaces for bodies that are not in John's position. Obviously we can apply a linear transformation T to bring any body K into John's position, but then a spherical section of $T(K)$ only translates back to an ellipsoidal section of the original body K. But there is an elegant way around this issue.

Recall that we can write any *ellipsoid* in the form $\mathcal{E} = \{\boldsymbol{x}\in\mathbb{R}^n \mid \sum_{i=1}^n \frac{\langle \boldsymbol{u}_i,\boldsymbol{x}\rangle^2}{\lambda_i^2} \leq 1\}$ where $\boldsymbol{u}_1,\dots,\boldsymbol{u}_n\in\mathbb{R}^n$ are orthonormal and $\lambda_1,\dots,\lambda_n > 0$ are the *axis lengths*. We will now see a surprising argument that any ellipsoid $\mathcal{E}\subseteq\mathbb{R}^n$ contains a slice of dimension $n/2$ where $\mathcal{E}$ is exactly a ball. Consider for example a 2-dimensional ellipsoid $\mathcal{E} = \{\boldsymbol{x}\in\mathbb{R}^2 \mid \frac{x_1^2}{\lambda_1^2} + \frac{x_2^2}{\lambda_2^2} \leq 1\}$ with $\lambda_1 < 1 < \lambda_2$. Then one can easily get a 1-dimensional slice that is a ball of radius 1.



Now the more general argument:

Lemma 5.5. *For any* $(2k-1)$*-dimensional ellipsoid* $\mathcal{E}$*, there is a* k*-dimensional subspace* L *so that* $\mathcal{E} \cap L$ *is a Euclidean ball.*

Proof. After translating, rotating and scaling we may assume that the ellipsoid is $\mathcal{E} = \{\boldsymbol{x} \in \mathbb{R}^{2k-1} \mid \sum_{i=1}^{2k-1} x_i^2/\lambda_i^2 \leq 1\}$ and the axis are sorted so that $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{2k-1}$ where the middle axis length is $\lambda_k = 1$. In order to generalize the 2-dimensional argument we form *pairs* of a long axis and a short axis, for example we can pair coordinate i with coordinate $2k-i$, so that $\lambda_i \leq 1 \leq \lambda_{2k-i}$. So we define a subspace

$$L := \left\{\boldsymbol{x} \in \mathbb{R}^{2k-1} \mid x_i\sqrt{\frac{1}{\lambda_i^2}-1} = x_{2k-i}\sqrt{1-\frac{1}{\lambda_{2k-i}^2}} \;\; \forall i = 1,\ldots,k-1\right\}$$

Squaring and rearranging the ith constraint in the definition of L will provide

$$\frac{x_i^2}{\lambda_i^2} + \frac{x_{2k-i}^2}{\lambda_{2k-i}^2} = x_i^2 + x_{2k-i}^2$$

Hence any point $\boldsymbol{x} \in L$ that lies on the boundary of $\mathcal{E}$ will have $\|\boldsymbol{x}\|_2^2 = 1$. □

Theorem 5.6. *Let* $K \subseteq \mathbb{R}^n$ *be a symmetric convex body and let* $0 < \varepsilon \leq \frac{1}{2}$*. Then there exists a subspace* V *of dimension* $k := \Theta(\frac{\varepsilon^2}{\log(1/\varepsilon)}\log(n))$ *so that* $K \cap V$ *is* $(1+\varepsilon)$*-spherical.*

Proof. Let T be a linear map so that $T(K)$ is in John's position. We know by combining Theorem 5.3 and Theorem 5.4 that for a random subspace V of dimension K, the section $T(K) \cap V$ is $(1+\varepsilon)$-spherical with good probability. Then $K \cap T^{-1}(V)$ is $(1+\varepsilon)$-ellipsoidal. Now use Lemma 5.5 to extract a subspace of half the dimension that is spherical. □

5.2 The critical dimension $k(K)$

For a convex symmetric body $K \subseteq \mathbb{R}^n$ and $0 < p < 1$we define the *critical dimension* $k_p(K)$ as the maximum value so that

$$\Pr_{\substack{V \subseteq \mathbb{R}^n:\\ \dim(V)=k_p(K)}}\Big[\frac{1}{2}\|\boldsymbol{x}\|_2 \cdot M(K) \leq \|\boldsymbol{x}\|_K \leq 2\|\boldsymbol{x}\|_2 \cdot M(K) \ \ \forall \boldsymbol{x} \in V\Big] \geq p$$

where V is a uniformly chosen subspace of dimension $k_p(K)$. We also abbreviate $k(K) := k_{1/2}(K)$. In other words, $k(K)$ is the maximum dimension so that most $k(K)$-dimensional subspaces are 2-spherical. We should add that in principle a section of a body could be spherical with respect to a different radius R, but as this has to hold for most subspaces V, we know that $R = (1 \pm o(1)) \cdot M(K)$ anyway.

Also note that the definition of critical dimension does not contain the error parameter ε — but this is also not necessary. There is always a $\Theta(\frac{\varepsilon^2}{\log(1/\varepsilon)} \cdot k(K))$-dimensional section of K that is $(1+\varepsilon)$-spherical. To see this, obtain first a $k(K)$-dimensional space V so that $K \cap V$ is 2-spherical. Then apply Theorem 5.3 again with parameter ε, using that $M(K \cap V) = \Theta(1)$.

Reinspecting the proof for Theorem 5.3 for $\varepsilon := \frac{1}{2}$ we see that we actually have a success probability of at least $1/2$ and so:

Theorem 5.7 (Dvoretzky-Milman)**.** *If $K \subseteq \mathbb{R}^n$ is centrally symmetric with $b \cdot B_2^n \subseteq K$. Then $k(K) \geq \Omega(n \cdot (M(K) \cdot b)^2)$.*

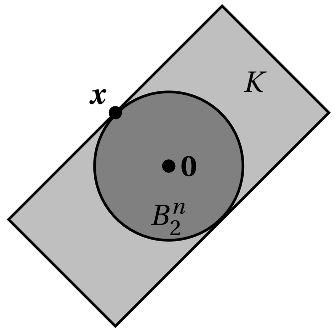
In particular every symmetric convex body K in John position has $k(K) \geq \Theta(\log n)$.

Somewhat surprisingly one can also prove that this is a tight bound for *every* symmetric body. In particular the critical dimension is also upper bounded in terms of $M(K)$ (where we need a little slack for the proof).

Theorem 5.8. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body so that B_2^n is the largest ball inside K. Then the critical dimension is $k_p(K) \leq O(n \cdot M(K)^2)$ where $p := 1 - \frac{1}{2n}$.*

Proof. We abbreviate $k := k_p(K)$ and $M := M(K)$ and assume for the sake of simplicity that n is a integer multiple of k. Fix any orthogonal subspaces $E^1, \ldots, E^{n/k} \subseteq \mathbb{R}^n$ with $\dim(E^i) = k$. Sample a random orthogonal transformation $U \in O(n)$ and consider the obtained random k-dimensional subspaces $U(E^1), \ldots, U(E^{n/k})$. By the union bound and the definition of $k_p(K)$ we know that with probability at least $1/2$, each section $K \cap U(E^i)$ is 2-spherical, meaning that $\frac{1}{2}M \cdot \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq 2\|\boldsymbol{x}\|_2 \cdot M$ for all $\boldsymbol{x} \in U(E^i)$ and all i. Next, fix one transformation U where this event happened.

As B_2^n is the biggest ball inside K, there is a contact point $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_K = 1$.



We can write $\boldsymbol{x} \in \mathbb{R}^n$ as $\boldsymbol{x} = \sum_{i=1}^{n/k} \boldsymbol{x}_i$ with $\boldsymbol{x}_i \in U(E^i)$. For this particular vector we have

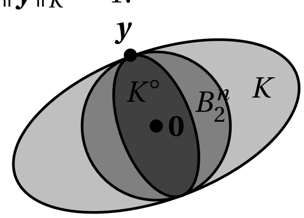
$$\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_K \overset{\text{triangle inequality}}{\leq} \sum_{i=1}^{n/k} \underbrace{\|\boldsymbol{x}_i\|_K}_{\leq 2M\|\boldsymbol{x}_i\|_2} \leq 2M \underbrace{\sum_{i=1}^{n/k} \|\boldsymbol{x}_i\|_2}_{\leq \sqrt{n/k}\|\boldsymbol{x}\|_2} \leq 2M\sqrt{\frac{n}{k}}\|\boldsymbol{x}\|_2$$

which can be rearranged to $k \leq 4M^2 n$. □

There is a useful corollary:

Corollary 5.9. Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body so that B_2^n is the largest ball inside K. Then $M(K) \geq \Theta(\frac{1}{\sqrt{n}})$.

Proof. The Corollary follows from Theorem 5.8 by rearranging $1 \leq k_p(K) \leq \Theta(n \cdot M(K)^2)$. However, one can also give a short self-contained argument. Take a vector $\boldsymbol{y}$ with $\|\boldsymbol{y}\|_K = 1 = \|\boldsymbol{y}\|_2$. Note that $1 = \langle \boldsymbol{y}, \boldsymbol{y} \rangle \leq \|\boldsymbol{y}\|_K \cdot \|\boldsymbol{y}\|_{K^\circ} = \|\boldsymbol{y}\|_{K^\circ}$ by Cauchy-Schwarz. But as $K^\circ \subseteq B_2^n$, the inequality must be tight and indeed the polar is touching K at y, i.e. $\|\boldsymbol{y}\|_{K^\circ} = 1$.



So we can lower bound

$$M(K) = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim S^{n-1}} [\|\boldsymbol{x}\|_K] \overset{\text{Cauchy-Schwarz}}{\geq} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim S^{n-1}} [|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|] \geq \Omega\Big(\frac{1}{\sqrt{n}}\Big).$$

□

The statement can be rephrased as follows:

Corollary 5.10. Every symmetric convex body $K \subseteq \mathbb{R}^n$ contains $\Theta(\frac{1}{\sqrt{n}\cdot M(K)}) \cdot B_2^n$.

Even if a symmetric convex body K is in John position, the largest 2-spherical subspace we can guarantee has dimension $\Theta(\log n)$. Quite surprisingly, one can prove that either K or the polar K° will have critical dimension $\Omega(\sqrt{n})$.

Theorem 5.11 (Figiel-Lindenstrauss-Milman [FLM77])**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body in John position. Then $k(K) \cdot k(K^\circ) \geq \Omega(n)$.*

Proof. Note that by assumption we have $B_2^n \subseteq K \subseteq \sqrt{n} B_2^n$. Next, the product of the M-values of K and its polar can be lower bounded as

$$M(K) \cdot M(K^\circ) \quad = \quad \mathop{\mathbb{E}}_{\boldsymbol{x} \in S^{n-1}}[\|\boldsymbol{x}\|_K] \cdot \mathop{\mathbb{E}}_{\boldsymbol{x} \in S^{n-1}}[\|\boldsymbol{x}\|_{K^\circ}] \overset{(*)}{\geq} \mathop{\mathbb{E}}_{\boldsymbol{x} \in S^{n-1}}\Big[\underbrace{\sqrt{\|\boldsymbol{x}\|_K \cdot \|\boldsymbol{x}\|_{K^\circ}}}_{\geq \|\boldsymbol{x}\|_2 = 1}\Big]^2 \overset{(**)}{\geq} 1$$

Here we use Hölder's Inequality in $(*)$ which says that for non-negative random variables X, Y one has $\mathbb{E}[X]\,\mathbb{E}[Y] \geq \mathbb{E}[\sqrt{XY}]^2$ (see Theorem 1.18 with $\lambda := \frac{1}{2}$). In $(**)$ we use Generalized Cauchy Schwarz in the form $|\langle \boldsymbol{x}, \boldsymbol{x}\rangle| \leq \|\boldsymbol{x}\|_K \cdot \|\boldsymbol{x}\|_{K^\circ}$. As $1 \cdot B_2^n \subseteq K$ and $\frac{1}{\sqrt{n}} \cdot B_2^n \subseteq K^\circ$ we obtain

$$k(K) \cdot k(K^\circ) \geq \Omega(n^2) \cdot \Big(M(K) \cdot 1\Big)^2 \cdot \Big(M(K^\circ) \cdot \frac{1}{\sqrt{n}}\Big)^2 \geq \Omega(n)$$

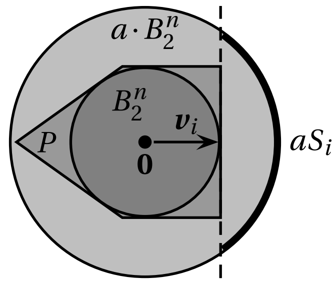
using the Dvoretzky-Milman Theorem (Theorem 5.7). □

5.3 Number of Faces and Vertices of Symmetric Polytopes.

Now we come to an application in combinatorics. For a *polytope* $P \subseteq \mathbb{R}^n$ let $f(P)$ be the *number of facets* and let $v(P)$ be the number of *vertices*. It turns out that almost spherical polytopes need to have a large number of facets. For example, if $B_2^n \subseteq P \subseteq n^{1/2-\varepsilon} \cdot B_2^n$ then the number of facets needs to be superpolynomial.

Lemma 5.12. *Let $P \subseteq \mathbb{R}^n$ be a polytope with m facets so that $B_2^n \subseteq P \subseteq a \cdot B_2^n$. Then $m \geq \frac{1}{4}\exp(\frac{n}{4a^2})$.*

Proof. W.l.o.g. push the inequalities defining P inwards until they touch B_2^n. Then we can we write $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{v}_i, \boldsymbol{x}\rangle \leq 1 \ \forall i \in [m]\}$ for some vectors $\boldsymbol{v}_i \in \mathbb{R}^n$ with $\|\boldsymbol{v}_i\|_2 = 1$.



Let $S_i := \{\boldsymbol{x} \in S^{n-1} \mid \langle \boldsymbol{x}, \boldsymbol{v}_i \rangle \geq \frac{1}{a}\}$. As $P \subseteq a \cdot B_2^n$, the union of the "arcs" $S_1, \ldots, S_m$ has to cover the sphere S^{n-1}. But by measure concentration, we know that the measure of each such cap is only $\Pr_{\boldsymbol{x} \sim S^{n-1}}[\langle \boldsymbol{v}_i, \boldsymbol{x} \rangle \geq \frac{1}{a}] \leq 4\exp(-\frac{n}{4a^2})$, see e.g. Lemma 3.4. Hence the claim follows. □

Now we can show that the critical dimension is at most logarithmic in the dimension. For example this implies that any polytope P with $n^{O(1)}$ any facets has a critical dimension of $k(P) \leq O(\log n)$.

Lemma 5.13. *For any polytope $P \subseteq \mathbb{R}^n$ one has $k(P) \leq O(\ln(f(P)))$.*

Proof. For $k := k(P)$, consider a subspace $U \subseteq \mathbb{R}^n$ so that $P \cap U$ is 2-spherical. Then by the previous Lemma and the fact that any facet in $P \cap U$ corresponds to a facet in P we get $f(P) \geq f(P \cap U) \geq \exp(\Theta(k))$. Taking logarithms then gives the claim. □

Now we can prove that any n-dimensional centrally symmetric polytope needs to have either $2^{\Omega(\sqrt{n})}$ many vertices or facets. Note that this claim does not holds for asymmetric polytopes — for example the simplex has $n+1$ facets and $n+1$ vertices.

Theorem 5.14 (Figiel-Lindenstrauss-Milman [FLM77])**.** *Let $P \subseteq \mathbb{R}^n$ be a centrally symmetric polytope. Then*

$$\ln(f(P)) \cdot \ln(v(P)) \geq \Omega(n).$$

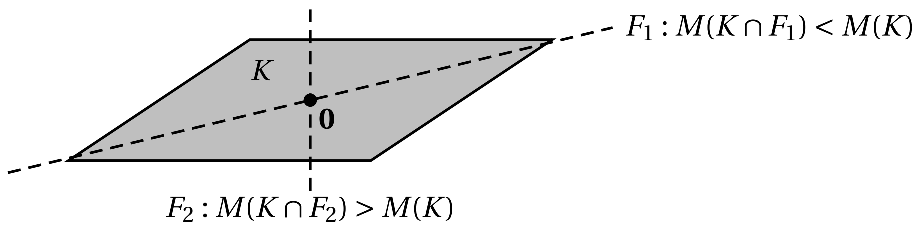
Proof. We can bring P into John position without changing the number of facets and vertices. Then

$$\ln(f(P)) \cdot \ln(v(P)) \overset{\text{polarity}}{=} \ln(f(P)) \cdot \ln(f(P^\circ)) \overset{\text{Lem. 5.13}}{\geq} \Omega(k(P) \cdot k(P^\circ)) \overset{\text{Thm. 5.11}}{\geq} \Omega(n)$$

□

5.4 The M-value for random subspaces

We would also like to discuss a result that is particularly useful in Chapter 7. First of all, if $F \subseteq \mathbb{R}^n$ is a proper subspace, then clearly $\mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_{K\cap F}] = \infty$. Hence the right definition of the M-value for a subspace is $M(K\cap F) := \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}\cap F}[\|\boldsymbol{x}\|_{K\cap F}]$. Note that as we intersect K with a subspace F, it is not trivial how the quantity $M(K\cap F) = \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}\cap F}[\|\boldsymbol{x}\|_{K\cap F}]$ relates to $M(K)$. Of course $\|\boldsymbol{x}\|_{K\cap F} = \|\boldsymbol{x}\|_K$ for every individual point $\boldsymbol{x} \in F$, but as we are averaging over a different set, the value $M(K\cap F)$ might be either larger or smaller than $M(K)$.



However we can prove that $M(K\cap F)$ is unlikely to increase more than a constant factor:

Lemma 5.15. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex set and let $k \in \{1,\dots,n\}$. Then*

$$\Pr[M(K\cap F) \geq C\cdot M(K)] \leq e^{-\Omega(k)}$$

where $F \subseteq \mathbb{R}^n$ is a uniform random k-dimensional subspace and C is a large enough constant.

Proof. We sample a uniform random k-dimensional subspace $F \subseteq \mathbb{R}^n$, by picking vectors $\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_k \sim S^{n-1}$ independently at random and letting $F := \text{span}\{\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_k\}$. It will be useful to abbreviate the random variable $X := \frac{1}{k}\sum_{i=1}^k \|\boldsymbol{\theta}_i\|_K$. Recall that we always have the rather weak concentration bound:

Claim I. *For any $\varepsilon > 0$ one has* $\Pr_{\boldsymbol{\theta}\sim S^{n-1}}[\|\boldsymbol{\theta}\|_K > (1+\varepsilon)M(K)] \leq e^{-\Theta(\varepsilon^2)}$.

Proof of Claim. After scaling we may assume that $B_2^n \subseteq K$ is the largest ball inside K. Then $\|\cdot\|_K$ is 1-Lipschitz and in Cor 5.9 we have seen that $M(K) \geq \Omega(\frac{1}{\sqrt{n}})$. The claim then follows from the usual concentration of 1-Lipschitz functions of S^{n-1}. □

This concentration is being amplified as the subspace F is spanned by k vectors:

Claim II. *One has* $\Pr_{\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_k\sim S^{n-1}}[X > C\cdot M(K)] \leq e^{-100k}$ *if C is a large enough constant.*

Proof of claim. We abbreviate $X_i := \|\boldsymbol{\theta}_i\|_K - M(K)$ where $\boldsymbol{\theta}_i \sim S^{n-1}$ so that $\mathbb{E}[X_i] = 0$. Note that $X - M(K) = \frac{1}{k}\sum_{i=1}^k X_i$ is also a mean-zero random variable. Combining Claim I and Lemma 3.15 we see that $\|X_i\|_{\psi_2} \leq O(1)\cdot M(K)$. Hence by

Lemma 3.17 we have $\|X - M(K)\|_{\psi_2} \leq O(\frac{1}{k}) \cdot (\sum_{i=1}^k \|X_i\|_{\psi_2}^2)^{1/2} \leq O(\frac{1}{\sqrt{k}}) \cdot M(K)$. Then again applying Lemma 3.15 and choosing $C > 0$ large enough we have

$$\Pr[X - M(K) > C \cdot M(K)] \leq e^{-100k}$$

which then gives the claim. $\square$

Claim III. *For fixed F one has $p_F := \Pr_{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_k \sim S^{n-1} \cap F}[X > \frac{1}{2} M(K \cap F)] \geq \Omega(\frac{1}{\sqrt{k}})$.*

Proof of claim. We can derive from Cor 5.9 also that for any $\boldsymbol{\theta} \in S^{n-1} \cap F$ one has $0 \leq \|\boldsymbol{\theta}\|_{K \cap F} \leq C\sqrt{k} M(K \cap F)$ — simply scale K until $B_2^n \cap F$ and $K \cap F$ touch in which case $0 \leq \|\boldsymbol{\theta}\|_{K \cap F} \leq 1 \leq C\sqrt{k} \cdot M(K \cap F)$. Next, observe that

$$M(K \cap F) = \mathop{\mathbb{E}}_{\boldsymbol{\theta} \sim S^{n-1} \cap F}[\|\boldsymbol{\theta}\|_K] = \mathbb{E}\Big[\frac{1}{k} \sum_{i=1}^k \|\boldsymbol{\theta}_i\|_K\Big] = \mathbb{E}[X \mid F]$$

Since we see that $0 \leq X \leq O(\sqrt{k}) \cdot M(K \cap F)$, by a Markov-type inequality one needs to have $\Pr[X \geq \frac{1}{2} M(K \cap F)] \geq \Omega(\frac{1}{\sqrt{k}})$. $\square$

Now we can put everything together. Recall that for arbitrary events $\mathcal{A}, \mathcal{B}$ one has the bound $\Pr[\mathcal{A}] = \frac{\Pr[\mathcal{A} \cap \mathcal{B}]}{\Pr[\mathcal{B} \mid \mathcal{A}]} \leq \frac{\Pr[\mathcal{B}]}{\Pr[\mathcal{B} \mid \mathcal{A}]}$. Hence

$$\begin{aligned} \Pr[M(K \cap F) \geq 2C \cdot M(K)] &\overset{\text{cond.prob}}{\leq} \frac{\Pr[X \geq C \cdot M(K)]}{\Pr[X \geq C \cdot M(K) \mid M(K \cap F) \geq 2C \cdot M(K)]} \\ &\overset{\text{Claim II+III}}{\leq} \frac{e^{-100k}}{\Theta(1/\sqrt{k})} \end{aligned}$$

Here we have implicitly used that conditioning on a subspace F, the distribution of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k$ are just independent uniform samples from $F \cap S^{n-1}$. $\square$

5.5 Euclidean Subspaces of ℓ_p^n

Recall that $B_p^n := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_p \leq 1\} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \sum_{i=1}^n |x_i|^p \leq 1\}$ is the ℓ_p-unit-ball. Also note that $1 \leq p \leq \infty$, the set B_p^n is convex.

B_p^n for $p = \frac{4}{3}$ sandwiched as $B_1^n \subseteq B_p^n \subseteq B_2^n$

B_p^n for $p = 4$ sandwiched as $B_2^n \subseteq B_p^n \subseteq B_\infty^n$

Depending on p, the balls B_p^n have very large almost Euclidean sections. For example the ball B_1^n of the $\|\cdot\|_1$-norm has critical dimension $\Theta(n)$,

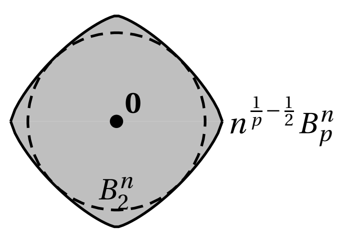
Theorem 5.16. *For any fixed $1 \leq p \leq 2$ one has $k(B_p^n) = \Theta(n)$.*

Proof. We first prove a variant of Hölder's Inequality:
Claim I. *For $0 < r < s$ one has $(\frac{1}{n}\sum_{i=1}^n |x_i|^r)^{1/r} \leq (\frac{1}{n}\sum_{i=1}^n |x_i|^s)^{1/s}$.*
Proof of claim. Define $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ with $f(x) := x^{s/r}$ which is a convex function. Moreover let X be the random variable that picks a uniform element from $|x_1|, \dots, |x_n|$. Then by Jensen's inequality

$$\Big(\frac{1}{n}\sum_{i=1}^n |x_i|^r\Big)^{s/r} = f(\mathbb{E}[X^r]) \leq \mathbb{E}[f(X^r)] = \frac{1}{n}\sum_{i=1}^n |x_i|^s$$

Applying $(\dots)^{1/s}$ to both sides then gives the claim. $\square$
Claim II. *For $1 \leq p \leq 2$ and all $\boldsymbol{x} \in \mathbb{R}^n$ one has $\big(\frac{1}{n}\big)^{1-\frac{1}{p}} \cdot \|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_p \leq n^{\frac{1}{p}-\frac{1}{2}} \cdot \|\boldsymbol{x}\|_2$.*
Proof of Claim. Applying Claim I with $r := p$ and $s := 2$ gives $\frac{1}{n^{1/p}}\|\boldsymbol{x}\|_p \leq \frac{1}{n^{1/2}}\|\boldsymbol{x}\|_2$. On the other hand, setting $r := 1$ and $s := p$ in Claim I we obtain $\frac{1}{n}\|\boldsymbol{x}\|_1 \leq \frac{1}{n^{1/p}}\|\boldsymbol{x}\|_p$. Combining both inequalities finishes Claim II. $\square$

Now we consider the body $K := n^{\frac{1}{p}-\frac{1}{2}} \cdot B_p^n$. Then the upper bound in Claim II guarantees that $\|\boldsymbol{x}\|_K \leq \|\boldsymbol{x}\|_2$ for all $\boldsymbol{x}$ and so $B_2^n \subseteq K$.



Moreover

$$M(K) \overset{\text{Def } M}{=} n^{\frac{1}{2}-\frac{1}{p}} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_p] \overset{\text{Claim II}}{\geq} n^{\frac{1}{2}-\frac{1}{p}} \cdot n^{\frac{1}{p}-1} \cdot \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_1]}_{=\Theta(\sqrt{n})} = \Theta(1)$$

Then the claim follows as $k(K) \geq \Omega(n \cdot M(K)^2) \geq \Omega(n)$. $\square$

It turns out that the hypercube $B_\infty^n = [-1,1]^n$ is one of the bodies minimizing the critical dimension:

Theorem 5.17. *One has $k(B_\infty^n) = \Theta(\log n)$.*

Proof. The lower bound follows from Theorem 5.7 as B_∞^n is in John position. For the upper bound consider a subspace $F \subseteq \mathbb{R}^n$ so that $B_\infty^n \cap F$ is 2-spherical. We can take a linear transformation T so that $P := T(B_\infty^n \cap F)$ is sandwiched as $B_2^k \subseteq P \subseteq 4B_2^k$. Every facet of P corresponds to a facet of B_∞^n and so $e^{\Theta(k)} \leq f(P) \leq f(B_\infty^n) \leq 2n$ by Lemma 5.13. Rearranging gives the claim. $\square$

We can also rather precisely characterize the critical dimension for B_p^n if $p > 2$. The intuition is that if p grows, then the norm $\|\cdot\|_p$ puts more and more weight on outliers and the norm is less similar to $\|\cdot\|_2$ which in turn means the maximum dimension of almost spherical subspaces will shrink.

Theorem 5.18. *For $2 \leq p < \infty$ one has $\Theta(n^{2/p}) \leq k(B_p^n) \leq \Theta(p \cdot n^{2/p})$.*

Proof. First note that we have $B_2^n \subseteq B_p^n$. Then for the lower bound it suffices to prove that $M(B_p^n) \geq n^{\frac{1}{p}-\frac{1}{2}}$ as then $n \cdot M(B_p^n)^2 \geq n \cdot (n^{\frac{1}{p}-\frac{1}{2}})^2 = n^{2/p}$. To get this estimate, we first use Claim I from Theorem 5.16 with parameters $0 < 2 \leq p$ to obtain $\|\boldsymbol{x}\|_p \geq n^{\frac{1}{p}-\frac{1}{2}} \|\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Then

$$M(B_p^n) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_p] \geq n^{\frac{1}{p}-\frac{1}{2}} \cdot \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_2]}_{=1} = n^{\frac{1}{p}-\frac{1}{2}}$$

For the upper bound we prove the stronger property that *any* 2-spherical section $B_p^n \cap V$ has $k := \dim(V) \leq O(p \cdot n^{2/p})$ — not just most random sections. Let $\boldsymbol{U} \in \mathbb{R}^{n\times k}$ be a matrix so that the column vectors $\boldsymbol{U}^1, \dots, \boldsymbol{U}^k$ form an orthonormal basis of the subspace V, that means

$$\frac{R}{2} \cdot \|\boldsymbol{y}\|_2 \leq \|\boldsymbol{U}\boldsymbol{y}\|_p \leq 2R \cdot \|\boldsymbol{y}\|_2 \quad (*)$$

for all $\boldsymbol{y} \in \mathbb{R}^k$ where R is the approximate radius of the spherical section. If we fix any index $i \in [n]$ and set $\boldsymbol{y} := \boldsymbol{U}_i$ in $(*)$ then

$$\|\boldsymbol{U}_i\|_2^2 \leq \Big(\sum_{i'=1}^n |\langle \boldsymbol{U}_{i'}, \boldsymbol{U}_i\rangle|^p\Big)^{1/p} = \|\boldsymbol{U}\boldsymbol{U}_i\|_p \overset{(*)}{\leq} 2R\|\boldsymbol{U}_i\|_2$$

and so $R \geq \frac{1}{2\|\boldsymbol{U}_i\|_2} \geq \frac{1}{2}$ as $\|\boldsymbol{U}_i\|_2 \leq 1$. Then taking random signs $\boldsymbol{y} \in \{-1,1\}^k$ gives

$$\begin{aligned}
(1/4)^p \cdot k^{p/2} &\overset{\substack{\|\boldsymbol{y}\|_2=\sqrt{k}\\ R\geq 1/2}}{\leq} (R/2)^p \mathop{\mathbb{E}}_{\boldsymbol{y}\in\{-1,1\}^k}\big[\|\boldsymbol{y}\|_2^p\big] \overset{(*)}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{y}\in\{-1,1\}^k}\Big[\big\|\boldsymbol{U}\boldsymbol{y}\big\|_p^p\Big] \\
&\overset{\text{Khintchine}}{\leq} (C\sqrt{p})^p \sum_{i=1}^n \Big(\underbrace{\|\boldsymbol{U}_i\|_2^2}_{\leq 1}\Big)^{p/2} \overset{(**)}{\leq} (C\sqrt{p})^p \cdot n
\end{aligned}$$

Here we apply Khintchine's inequality n times to each coordinate contribution $\mathbb{E}_{\boldsymbol{y}\in\{-1,1\}^k}[|\langle \boldsymbol{U}_i,\boldsymbol{y}\rangle|^p]$. Note that a matrix $\boldsymbol{U}$ whose columns form an orthonormal basis (even for a subspace) has the property that each row $i\in[n]$ satisfies $\|\boldsymbol{U}_i\|_2^2\leq 1$. Rearranging gives $k\leq O(p\cdot n^{2/p})$. $\square$

Note that this bound does not give tight estimates for the whole range of parameters. While for the border case $p=2$ it gives the obvious bound of $\Theta(n)\leq k(B_2^n)\leq\Theta(n)$, there is a growing gap for larger p. Recall that the parameter range $p:=\log_2(n)$ approximates the cube well enough as $\frac{1}{2}B_\infty^n\subseteq B_{\log_2(n)}^n\subseteq B_\infty^n$. For this parameter, the theorem gives the bound of $\Theta(1)\leq k(B_{\log_2(n)}^n)\leq\Theta(\log n)$ and here we know that the upper bound is tight and not the lower bound. In fact, [PVZ17] prove that $k(B_p^n)=\Theta(pn^{2/p})$ for $2\leq p\leq\Theta(\log n)$.

5.6 Kashin's Theorem

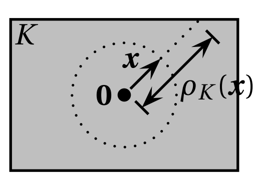
We have seen that for a symmetric body $K\supseteq B_2^n$, the dimension of almost spherical sections is controlled by the quantity $M(K)$. However, a different quantity that could be used is simply the *volume* of K. For the sake of argument suppose that K is in John's position and we write $\mathrm{Vol}_n(K)=\alpha^n\cdot\mathrm{Vol}_n(B_2^n)$. Then we know from John's theorem that $1\leq\alpha\leq\sqrt{n}$. On the other hand it may very well be that α is a lot smaller than this pathological upper bound, say $\alpha=O(1)$. In this case we can see that the body has $n/2$-dimensional sections that are $O(1)$-spherical.

Theorem 5.19 (Volume Ratio Theorem - Kashin, Szarek, Tomczak-Jaegermann [Sza77, STJ80, Kas77])**.** *Let $K\subseteq\mathbb{R}^n$ be a centrally symmetric convex body with $B_2^n\subseteq K$ and $\mathrm{Vol}_n(K)=\alpha^n\mathrm{Vol}_n(B_2^n)$ for some $\alpha\geq 1$. For $k\in\{1,\dots,n\}$, a random k-dimensional subspace $F\subseteq\mathbb{R}^n$ has*

$$B_2^n\cap F\subseteq K\cap F\subseteq(C\alpha)^{\frac{n}{n-k}}(B_2^n\cap F)$$

with probability at least $1-2^{-n}$ where C is an absolute constant.

Proof. Recall that for a unit vector $\boldsymbol{x}\in S^{n-1}$, the quantity $\rho_K(\boldsymbol{x}):=\frac{1}{\|\boldsymbol{x}\|_K}\geq 1$ denotes the *radius* of K in direction $\boldsymbol{x}$.



Writing the volume of K in polar coordinates we obtain

$$\mathop{\mathbb{E}}_{F}\Big[\mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}\cap F}\big[\rho_K(\boldsymbol{x})^n\big]\Big] = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}\big[\rho_K(\boldsymbol{x})^n\big] \overset{\text{Lem 1.46}}{=} \frac{\mathrm{Vol}_n(K)}{\mathrm{Vol}_n(B_2^n)} = \alpha^n$$

Then by Markov's inequality with probability $1-2^{-n}$, a random subspace $F \subseteq \mathbb{R}^n$ will satisfy

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}\cap F}\big[\rho_K(\boldsymbol{x})^n\big] \leq (2\alpha)^n \quad (*)$$

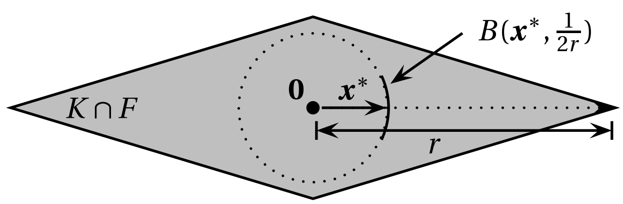
So it remains to prove the following:

Claim. *Any k-dimensional subspace F satisfying $(*)$ has $B_2^n \cap F \subseteq K \cap F \subseteq (O(\alpha))^{\frac{n}{n-k}} \cdot (B_2^n \cap F)$.*

Proof of claim. Let $\boldsymbol{x}^* \in S^{n-1}\cap F$ be point maximizing the radius $\rho_K(\boldsymbol{x})$ and let us abbreviate $r := \rho_K(\boldsymbol{x}^*)$ as that radius. We need to argue that there is a even significant fraction of points with high radius. Let $B(\boldsymbol{x}^*, \frac{1}{2r}) := \{\boldsymbol{y} \in S^{n-1}\cap F \mid \|\boldsymbol{x}^* - \boldsymbol{y}\|_2 \leq \frac{1}{2r}\}$ be the spherical cap of radius $\frac{1}{2r}$. Then

$$\|\boldsymbol{y}\|_K \leq \|\boldsymbol{x}\|_K + \|\boldsymbol{x}-\boldsymbol{y}\|_K \overset{B_2^n\subseteq K}{\leq} \frac{1}{r} + \|\boldsymbol{x}-\boldsymbol{y}\|_2 \leq \frac{2}{r}$$

which means that the radius is $\rho_K(\boldsymbol{y}) \geq \frac{r}{2}$ for all points $\boldsymbol{y} \in B(\boldsymbol{x}^*, \frac{1}{2r})$ in the spherical cap. On the other hand we have discussed earlier that the volume of a k-dimensional spherical cap satisfies[1] $\sigma_F(B(\boldsymbol{x}^*, \frac{1}{2r})) \geq (\frac{1}{8r})^k$.



On the other hand, we know that

$$\mathop{\Pr}_{\boldsymbol{x}\sim S^{n-1}\cap F}\Big[\rho_K(\boldsymbol{x}) > \frac{r}{2}\Big]\cdot (r/2)^n \leq \mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}\cap F}\big[\rho_K(\boldsymbol{x})^n\big] \leq (2\alpha)^n$$

In particular, the set $A_{r/2} := \{\boldsymbol{x} \in S^{n-1}\cap F \mid \rho_K(\boldsymbol{x}) \geq r/2\}$ has measure $\sigma_F(A_{r/2}) \leq (\frac{4\alpha}{r})^n$. Then

$$\Big(\frac{1}{8r}\Big)^k \leq \sigma_F\Big(B\Big(\boldsymbol{x}^*, \frac{1}{2r}\Big)\Big) \overset{B(\boldsymbol{x}^*,\frac{1}{2r})\subseteq A_{r/2}}{\leq} \sigma_F(A_r) \leq \Big(\frac{2\alpha}{r}\Big)^n$$

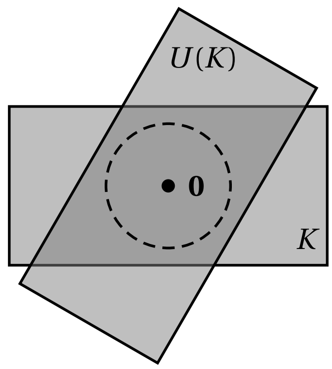
Then rearranging gives $r \leq (8^k \cdot (2\alpha)^n)^{1/(n-k)}$ which then proves the claim. $\square$

[1] In Lemma 5.1 we have seen that S^{k-1} can be covered with at most $(\frac{4}{\varepsilon})^k$ many spherical caps $B(\boldsymbol{y},\varepsilon)$ and hence each individual cap must have a measure of $\sigma(B(\boldsymbol{y},\varepsilon)) \geq (\frac{\varepsilon}{4})^k$. The claim then follows if we set $\varepsilon := \frac{1}{2r}$.

Next, we prove that for a symmetric body $K \supseteq B_2^n$ whose volume is not much bigger than the volume of the ball, the intersection with a random rotation $U(K)$ is close to B_2^n. The intuition is that even if K has very long spikes, they will be few so that the intersection cuts them off:

Theorem 5.20. *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex body with $B_2^n \subseteq K$ and $Vol_n(K) = \alpha^n Vol(B_2^n)$ for some $\alpha > 1$. Then there is an orthogonal transformation $U \in O(n)$ so that*

$$B_2^n \subseteq K \cap U(K) \subseteq O(\alpha^2) \cdot B_2^n$$



Proof. For the sake of simplicity suppose n is even. We have seen in the last theorem that a random subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) = \frac{n}{2}$ has $B_2^n \cap F \subseteq K \cap F \subseteq O(\alpha^2) \cdot (B_2^n \cap F)$ with high probability. Note that also the orthogonal complement $F^\perp$ is a random subspace with $\dim(F^\perp) = \frac{n}{2}$. Let $P_F : \mathbb{R}^n \to F$ be the orthogonal projection into the subspace F. Then we define $U(\boldsymbol{x}) := P_F(\boldsymbol{x}) - P_{F^\perp}(\boldsymbol{x})$. Note that if we write $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ with $\boldsymbol{x}_1 \in F$ and $\boldsymbol{x}_2 \in F^\perp$, then we have $U(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \boldsymbol{x}_1 - \boldsymbol{x}_2$. In particular from this we see that U is indeed an orthogonal transformation and moreover, it has the useful property that $U^{-1} = U$. Geometrically speaking, U is the *reflection* w.r.t. the subspace $F^\perp$. It remains to prove that $\|\boldsymbol{x}\|_2 \geq \|\boldsymbol{x}\|_{K \cap U(K)} \geq \Omega(\frac{1}{\alpha^2}) \cdot \|\boldsymbol{x}\|_2 \; \forall \boldsymbol{x} \in \mathbb{R}^n$. As $B_2^n \subseteq K$ we also have $B_2^n \subseteq U(K)$. Then

$$\begin{aligned}
\|\boldsymbol{x}\|_2 &\overset{B_2^n \subseteq K \cap U(K)}{\geq} \|\boldsymbol{x}\|_{K \cap U(K)} = \max\{\|\boldsymbol{x}\|_K, \|\boldsymbol{x}\|_{U(K)}\} \overset{(*)}{=} \max\{\|\boldsymbol{x}\|_K, \|U(\boldsymbol{x})\|_K\} \\
&\overset{\text{Def } U}{=} \max\{\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_K, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_K\} \overset{(**)}{\geq} \frac{1}{2}(\|\boldsymbol{x}_1\|_K + \|\boldsymbol{x}_2\|_K) \\
&\overset{(***)}{\geq} \Omega\Big(\frac{1}{\alpha^2}\Big) \cdot (\|\boldsymbol{x}_1\|_2 + \|\boldsymbol{x}_2\|_2) \geq \Omega\Big(\frac{1}{\alpha^2}\Big) \cdot \|\boldsymbol{x}\|_2
\end{aligned}$$

In $(*)$ we use that $\boldsymbol{x} = \lambda U(\boldsymbol{y}) \Leftrightarrow U(\boldsymbol{x}) = U^{-1}(\boldsymbol{x}) = \lambda \boldsymbol{y}$ and so $\|\boldsymbol{x}\|_{U(K)} = \|U(\boldsymbol{x})\|_K$. A argument for $(**)$ is the following: suppose that $r := \max\{\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_K, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_K\}$ meaning that $\pm \boldsymbol{x}_1 \pm \boldsymbol{x}_2 \in rK$. Then by convexity $\boldsymbol{x}_1, \boldsymbol{x}_2 \in rK$ which implies that $\max\{\|\boldsymbol{x}_1\|_K, \|\boldsymbol{x}_2\|_K\} \leq \max\{\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_K + \|\boldsymbol{x}_1 - \boldsymbol{x}\|_K\}$.

For $(***)$ we used that $K \cap F \subseteq O(\alpha^2) B_2^n \cap F$ and hence $\|\boldsymbol{x}_1\|_2 \geq \Omega(\frac{1}{\alpha^2}) \|\boldsymbol{x}_1\|_K$ — the same holds for $\boldsymbol{x}_2$. □

A particularly useful and instructive case is if $K = B_1^n$. We can then prove that there is a matrix $\boldsymbol{A} \in \mathbb{R}^{2n \times n}$ so that $\|\boldsymbol{A}\boldsymbol{x}\|_1 = \Theta(\|\boldsymbol{x}\|_2)$ for every $\boldsymbol{x} \in \mathbb{R}^n$:

Theorem 5.21. *There are vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{2n} \in S^{n-1}$ so that $\sum_{j=1}^{2n} |\langle \boldsymbol{x}, \boldsymbol{y}_j \rangle| = \Theta(\sqrt{n}) \cdot \|\boldsymbol{x}\|_2$ for every $\boldsymbol{x} \in \mathbb{R}^n$.*

Proof. We know that $B_2^n \subseteq \sqrt{n} B_1^n$ and $\mathrm{Vol}_n(\sqrt{n} B_1^n) \leq O(1)^n \cdot \mathrm{Vol}_n(B_2^n)$ hence by the previous theorem there is an orthogonal transformation U so that

$$\sum_{i=1}^n |\langle \boldsymbol{e}_i, \boldsymbol{x} \rangle| + \sum_{i=1}^n |\langle \boldsymbol{U}_i, \boldsymbol{x} \rangle| = \|\boldsymbol{x}\|_1 + \|\boldsymbol{U}\boldsymbol{x}\|_1 = \Theta(\sqrt{n}) \cdot \|\boldsymbol{x}\|_2$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. □

A natural question is whether one can obtain a $(1+\varepsilon)$-spherical section by allowing more than one random orthogonal transformation. Indeed this is true. For space reasons, we skip the proof which again is a concentration argument:

Theorem 5.22 (Bourgain-Lindenstrauss-Milman)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body so that B_2^n is the largest radius ball contained in K. If we sample independent random orthogonal transformations $U_1, \ldots, U_t$ where $t = \Theta(\frac{1}{\varepsilon^2 M(K)^2})$, then with high probability one has*

$$\frac{M(K)}{1+\varepsilon} \leq \mathop{\mathbb{E}}_{j \sim [t]} \left[\|U_j(\boldsymbol{x})\|_K \right] \leq (1+\varepsilon) \cdot M(K) \quad \forall \boldsymbol{x} \in S^{n-1}$$

5.7 Diameter of random projections

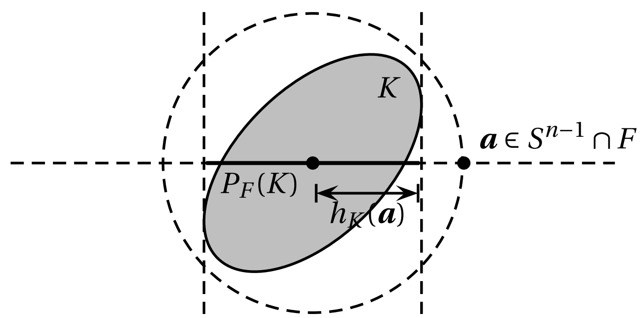
We will prove a result about the diameter of random projections of convex symmetric sets. This lemma will be useful later in Chapter 7. For a symmetric convex body $K \subseteq B_2^n$, we define the *radius* as $\mathrm{radius}(K) := \max\{\|\boldsymbol{x}\|_2 \mid \boldsymbol{x} \in K\}$. Phrased differently, the radius is the minimum number r so that $K \subseteq r B_2^n$. We claim that a proxy for the radius of "round" convex sets should be the quantity $M(K^\circ) = \frac{1}{2} w(K)$. To justify this, suppose that $K = r B_2^n$. Then $M(K^\circ) = \mathbb{E}_{\boldsymbol{x} \in S^{n-1}}[\|\boldsymbol{x}\|_{K^\circ}] = \mathbb{E}_{\boldsymbol{x} \in S^{n-1}}[\max\{\langle \boldsymbol{y}, \boldsymbol{x} \rangle : \boldsymbol{y} \in K\}] = r = \mathrm{radius}(K)$. Of course in general, K could have long spikes that do not contribute much to the average value $M(K^\circ)$. But it turns out that random projections of symmetic convex sets are getting close to attaining this value. Recall that for a subspace $F \subseteq \mathbb{R}^n$, $P_F : \mathbb{R}^n \to F$ denotes the *orthogonal projection* into F.

Lemma 5.23. *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex set with $K \subseteq rB_2^n$ and let $k \in \{1,\dots,n\}$. Then*

$$\Pr_{\dim(F)=k}\left[P_F(K) \subseteq O(1)\cdot\max\left\{M(K^\circ), r\sqrt{\tfrac{k}{n}}\right\}\right] \geq 1 - e^{-\Omega(k)}$$

where F is a uniform random k-dimensional subspace.

Proof. After rescaling we may assume that $r = 1$. Recall that $h_K(\boldsymbol{a}) := \max\{\langle \boldsymbol{a}, \boldsymbol{x}\rangle : \boldsymbol{x} \in K\} = \|\boldsymbol{a}\|_{K^\circ}$ is the *support function* of K. We have $K \subseteq \rho B_2^n$ iff $h_K(\boldsymbol{a}) \leq \rho$ for every $\boldsymbol{a} \in S^{n-1}$. For the projection $P_F(K)$, the radius is then $\max_{\boldsymbol{a}\in S^{n-1}\cap F} h_K(\boldsymbol{a})$. Also the projection of K is convex and symmetric, hence from previous arguments we know that if $\mathcal{N}$ is a $\frac{1}{2}$-net of $F \cap S^{n-1}$, then the radius of $P_F(K)$ is $\Theta(1)\cdot\max_{\boldsymbol{a}\in\mathcal{N}} h_K(\boldsymbol{a})$.



So, we fix a $\frac{1}{2}$-net $\mathcal{N}$ of $\text{span}\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_k\}$ of size $|\mathcal{N}| \leq 5^k$. Let U be a uniformly random orthogonal transformation. As we did before, we set $F := \text{span}\{U(\boldsymbol{e}_1),\dots,U(\boldsymbol{e}_k)\}$ as the uniformly sampled subspace. As $K \subseteq B_2^n$, the function h_K is 1-Lipschitz. Moreover, $\mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[h_K(\boldsymbol{x})] = \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_{K^\circ}] = M(K^\circ)$. Then by Lemma 3.5,

$$\Pr_{\boldsymbol{x}\sim S^{n-1}}\left[h_K(\boldsymbol{x}) > 10000\cdot\left(M(K^\circ) + \sqrt{\tfrac{k}{n}}\right)\right] \leq 64\exp\left(-10n\cdot\left(\sqrt{\tfrac{k}{n}}\right)^2\right) = 64\exp(-10k)$$

In particular we can apply the union bound over all points in the $\frac{1}{2}$-net to obtain

$$\Pr_U\left[h_K(U(\boldsymbol{x})) \leq 10000\cdot\left(M(K^\circ) + \sqrt{\tfrac{k}{n}}\right)\ \forall \boldsymbol{x} \in \mathcal{N}\right] \geq 1 - 64e^{-4k}$$

If the latter event happens, the radius of $P_F(K)$ is indeed bounded as claimed. $\square$

5.8 Exercises

Exercise 5.1.

Let $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m \in S^{n-1}$ and let $c_1,\dots,c_m \geq 0$ be coefficients so that $\sum_{i=1}^m c_i \boldsymbol{a}_i \boldsymbol{a}_i^T = \boldsymbol{I}_n$. Prove that $\mathbb{E}_{\boldsymbol{\theta}\sim S^{n-1}}[\max_{i=1,\dots,m} |\langle \boldsymbol{a}_i, \boldsymbol{\theta}\rangle|] \geq \Omega(\frac{\sqrt{\log(n)}}{\sqrt{n}})$.
Hint. Use polarity and a result from this chapter.

Exercise 5.2.
Let $N \geq 2$. We say that points $\boldsymbol{a}_1,\dots,\boldsymbol{a}_N \in S^{n-1}$ are *δ-separated* if $\|\boldsymbol{a}_i - \boldsymbol{a}_j\|_2 \geq \delta$ for all $i \neq j$. In the following you may use the following fact without proof[2]: *For any $\frac{1}{2}$-separated points $\boldsymbol{a}_1,\dots,\boldsymbol{a}_N \in S^{n-1}$ one has $\mathbb{E}_{\boldsymbol{\theta}\sim S^{n-1}}[\max_{i=1,\dots,N} |\langle \boldsymbol{a}_i, \boldsymbol{\theta}\rangle|] = \Theta(\sqrt{\log(N)/n})$.* Consider a polytope $K := \{\boldsymbol{x} \in \mathbb{R}^n \mid |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| \leq 1 \ \forall i \in [N]\}$ where the points $\boldsymbol{a}_1,\dots,\boldsymbol{a}_N \in S^{n-1}$ are $\frac{1}{2}$-separated.

(i) Prove that $k(K) \geq \Omega(\log(N))$.
(ii) Prove that $k(K) \leq O(\log(N))$.

[2] We will prove this later in Chapter 9.

Chapter 6

Pisier's Inequality and the MM°-estimate

We have seen that for a convex body $K \subseteq \mathbb{R}^n$ there is always an affine linear transformation T so that $T(K)$ approximates a ball up to a factor of n and this is best possible in general, if we ask for an inclusion property of the form $B_2^n \subseteq T(K) \subseteq nB_2^n$.

However, it turns out that there are "milder" properties where a convex body can be drastically better approximated by a ball. Recall that $w(K) := \mathbb{E}_{\boldsymbol{x}\in S^{n-1}}[w_K(\boldsymbol{x})]$ gives the *mean width* of a body. Suppose that we scale K so that $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(B_2^n)$. Then from *Urysohn's Inequality* we know that $w(K) \geq w(B_2^n) = 2$. On the other hand, $w(K)$ can be arbitrarily large for example if K is long and skinny in some direction. The main conclusion of this chapter will be a surprisignly strong upper bound:

> *For any convex body $K \subseteq \mathbb{R}^n$, there is a linear transformation T so that $\mathrm{Vol}_n(T(K)) = \mathrm{Vol}_n(B_2^n)$ and $w(T(K)) \leq O(\log n)$.*

If K is symmetric and T is the linear map so that $T(K)$ is in John position, then we know that $M(T(K)) \cdot M(T(K)^\circ) \leq \sqrt{n}$. This can be drastically improved:

> *MM°-estimate: For any symmetric convex body $K \subseteq \mathbb{R}^n$, there is a linear transformation T so that $M(T(K)) \cdot M(T(K)^\circ) \leq O(\log n)$.*

As $w(K) = 2M(K^\circ)$ this is equivalent to $w(T(K)) \cdot w(T(K)^\circ) \leq O(\log n)$.

6.1 Pisier's inequality

We will need to develop some machinery to achieve this result. In particular we will prove *Pisier's Inequality* which can be stated as follows: Consider the quantity $\mathbb{E}_{\boldsymbol{x}\in\{-1,1\}^n}[\|f(\boldsymbol{x})\|_K^2]^{1/2}$ for an arbitrary function $f:\{-1,1\}^n \to \mathbb{R}^m$. If we replace $f(\boldsymbol{x})$ by its *linear part* then this quantity can only increase by a factor of at most $O(\log m)$. This holds true for any norm $\|\cdot\|_K$. To derive this result, we revisit the concept of *Fourier analysis.*

6.1.1 Fourier analysis and the Rademacher Projection

For the remainder of this chapter, we will use μ_n as the uniform distribution on $\{-1,1\}^n$. For functions $f,g:\{-1,1\}^n \to \mathbb{R}$, we abbreviate $\langle f,g\rangle_{\mu_n} := \mathbb{E}_{\boldsymbol{x}\sim\{-1,1\}^n}[f(\boldsymbol{x})\cdot g(\boldsymbol{x})]$. In the literature, $\langle\cdot,\cdot\rangle_{\mu_n}$ is often called the *expectation inner product.* As the name suggests, this is indeed an inner product — we will get back to that in a bit. For any subset $A\subseteq[n]$ we define the *Walsh function* $w_A:\{-1,1\}^n \to \{-1,1\}$ by

$$w_A(\boldsymbol{x}) := \prod_{i\in A} x_i$$

Note that $w_\emptyset \equiv 1$ is the constant-1 function. The following is a rather basic fact in Fourier analysis:

Lemma 6.1. *The Walsh functions have the following property:*

(1) For $A,B\subseteq[n]$ one has

$$\langle w_A, w_B\rangle_{\mu_n} = \begin{cases} 1 & \text{if } A = B \\ 0 & \text{if } A \neq B \end{cases}$$

That means $\{w_A\}_{A\subseteq[n]}$ forms an orthonormal basis of the 2^n-dimensional vector space of functions $f:\{-1,1\}^n \to \mathbb{R}$.

(2) For $\boldsymbol{x},\boldsymbol{y}\in\{-1,1\}^n$ one has

$$\mathop{\mathbb{E}}_{A\subseteq[n]}[w_A(\boldsymbol{x})\cdot w_A(\boldsymbol{y})] = \begin{cases} 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} \\ 0 & \text{if } \boldsymbol{x} \neq \boldsymbol{y} \end{cases}$$

(3) Every function $f:\{-1,1\}^n \to \mathbb{R}^m$ can be written uniquely in the form $f(\boldsymbol{x}) = \sum_{A\subseteq[n]} w_A(\boldsymbol{x})\cdot \boldsymbol{y}_A$ for some $\boldsymbol{y}_A\in\mathbb{R}^m$. We call the vector $\hat{\boldsymbol{f}}_A := \boldsymbol{y}_A$ the A-th Fourier coefficient *of f.*

Proof. For sets A,B, let $A\triangle B$ be the *symmetric difference.* Then $w_A(\boldsymbol{x})\cdot w_B(\boldsymbol{x}) = \prod_{i\in A} x_i \cdot \prod_{i\in B} x_i = w_{A\triangle B}(\boldsymbol{x})$. For $S := A\Delta B$ one has

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}[w_S(\boldsymbol{x})] = \prod_{i\in S} \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}[x_i]}_{=0} = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{if } S \neq \emptyset \end{cases}$$

which shows (1). For (2) note that $w_A(\boldsymbol{x}) \cdot w_A(\boldsymbol{y}) = w_A(\boldsymbol{z})$ where $\boldsymbol{z} := \boldsymbol{x} \odot \boldsymbol{y}$ is given by the coordinate-wise multiplication. Then

$$\mathop{\mathbb{E}}_{A\subseteq[n]}[w_A(\boldsymbol{z})] = \mathop{\mathbb{E}}_{A\subseteq[n]}\Big[\prod_{i=1}^{n}\begin{Bmatrix} z_i & \text{if } i \in A \\ 1 & \text{if } i \notin A \end{Bmatrix}\Big] = \prod_{i=1}^{n} \underbrace{\mathop{\mathbb{E}}_{A\subseteq[n]}\Big[\begin{Bmatrix} z_i & \text{if } i \in A \\ 1 & \text{if } i \notin A \end{Bmatrix}\Big]}_{=0 \text{ if } z_i=-1, =1 \text{ if } z_i=1} = \begin{cases} 1 & \text{if } \boldsymbol{z} = (1,\dots,1) \\ 0 & \text{if } \boldsymbol{z} \neq (1,\dots,1) \end{cases}$$

For (3), consider first the case that $m = 1$. As the Walsh function form an orthonormal basis, there will clearly be one unique choice for y_A and that choice has to be the inner product with the Walsh function, i.e. $y_A = \langle f, w_A\rangle_{\mu_n}$. For $m > 1$ we apply the same reasoning coordinate wise — in fact $\boldsymbol{y}_A = (\langle f_1, w_A\rangle_{\mu_n}, \dots, \langle f_m, w_A\rangle_{\mu_n})$ where we denote $f_i : \{-1,1\}^n \to \mathbb{R}$ as the ith coordinate of f. □

Apart from an inner product, we also want to define *norms* for such functions. If $K \subseteq \mathbb{R}^m$ is a symmetric convex body and $f : \{-1,1\}^n \to \mathbb{R}^m$, then for $q \geq 1$, we can define a norm

$$\|f\|_{\mu_n \to \|\cdot\|_K^q} := \Big(\mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\big[\|f(\boldsymbol{x})\|_K^q\big]\Big)^{1/q}$$

In particular, the remainder of this section we will work towards understanding the behavior of the norm of a function if we "wipe out" the non linear part of the function. First we define formally what we mean with the "wipe out" part:

Definition 6.2. Let $f : \{-1,1\}^n \to \mathbb{R}^m$ be a function. Then the *Rademacher projection* $\mathrm{Rad}_n f : \{-1,1\}^n \to \mathbb{R}^m$ is defined by

$$\mathrm{Rad}_n f(\boldsymbol{x}) := \sum_{i=1}^{n} x_i \hat{\boldsymbol{f}}_{\{i\}}$$

Note that the Rademacher projection is a *linear operator* that maps a function f to the function just consisting of its linear parts.

For a symmetric convex body $K \subseteq \mathbb{R}^m$ we define the *convexity constant* as

$$\kappa(K) := \sup_{n,f}\Big\{\|\mathrm{Rad}_n f\|_{\mu_n \to \|\cdot\|_K^2} \mid f : \{-1,1\}^n \to \mathbb{R}^m \text{ with } \|f\|_{\mu_n \to \|\cdot\|_K^2} \leq 1\Big\}$$

Phrased differently, the convexity constant is the smallest value so that for any $n \in \mathbb{N}$ and any function $f : \{-1,1\}^n \to \mathbb{R}^m$ one has

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\Big[\Big\|\sum_{i=1}^{n} x_i \hat{\boldsymbol{f}}_{\{i\}}\Big\|_K^2\Big]^{1/2} \leq \kappa(K) \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^n}\Big[\big\|f(\boldsymbol{x})\big\|_k^2\Big]^{1/2}$$

Note that we make no restriction on f whatsoever. For example f might not even have any linear parts, i.e. it could be $\hat{\boldsymbol{f}}_{\{i\}} = \mathbf{0}$ for all i. In this case, $\|\mathrm{Rad}_n f\|_{\mu_n \to \|\cdot\|_K^2} = 0$ while $\|f\|_{\mu_n \to \|\cdot\|_K^2}$ could be arbitrarily large. In particular, this means that the norm of the function can go arbitrarily *down* when wiping out non-linear parts — however, we will see that the value cannot go *up* arbitrarily.

Next, we need another standard tool in Fourier analysis, which is the *convolution* of functions.

Definition 6.3. For functions $f : \{-1,1\}^n \to \mathbb{R}^m$ and $g : \{-1,1\}^n \to \mathbb{R}$ we define the *convolution* as the function $f * g : \{-1,1\}^n \to \mathbb{R}^m$ with

$$(f * g)(\boldsymbol{x}) := \mathop{\mathbb{E}}_{\boldsymbol{y} \in \{-1,1\}^n} \left[f(\boldsymbol{x} \odot \boldsymbol{y}) \cdot g(\boldsymbol{y}) \right] \quad \forall \boldsymbol{x} \in \{-1,1\}^n$$

Here $(\boldsymbol{x} \odot \boldsymbol{y})_i := x_i \cdot y_i$ is the coordinate-wise product. Note that the convolution is defined in an asymmetric way as the second function g is not vector-valued. We show that the Fourier coefficients of the convolution are simply the (scalar-)products of the Fourier coefficients of f and g. Note that $\hat{\boldsymbol{f}}_A \in \mathbb{R}^m$ is a vector and $\hat{g}_A \in \mathbb{R}$ is a scalar.

Lemma 6.4. *For functions* $f : \{-1,1\}^n \to \mathbb{R}^m$ *and* $g : \{-1,1\}^n \to \mathbb{R}$ *one has*

$$\widehat{\boldsymbol{f} * \boldsymbol{g}}_A = \hat{\boldsymbol{f}}_A \cdot \hat{g}_A \quad \forall A \subseteq [n]$$

Proof. The proof follows from writing out the definition of convolution and swapping the order of the expectations properly:

$$\begin{aligned}
\widehat{(\boldsymbol{f} * \boldsymbol{g})}_A &\stackrel{\text{Def}\,\widehat{}}{=} \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} \left[w_A(\boldsymbol{x}) \cdot (f * g)(\boldsymbol{x}) \right] \\
&\stackrel{\text{Def}\,*}{=} \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} \left[\mathop{\mathbb{E}}_{\boldsymbol{y} \in \{-1,1\}^n} \left[w_A(\boldsymbol{x}) \cdot f(\boldsymbol{x} \odot \boldsymbol{y}) \cdot g(\boldsymbol{y}) \right] \right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{y} \in \{-1,1\}^n} \left[\mathop{\mathbb{E}}_{\boldsymbol{z} \in \{-1,1\}^n} \left[w_A(\boldsymbol{z} \odot \boldsymbol{y}) \cdot f(\boldsymbol{z}) \cdot f(\boldsymbol{y}) \right] \right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{z} \in \{-1,1\}^n} \left[w_A(\boldsymbol{z}) \cdot f(\boldsymbol{z}) \right] \cdot \mathop{\mathbb{E}}_{\boldsymbol{y} \in \{-1,1\}^n} \left[w_A(\boldsymbol{y}) \cdot g(\boldsymbol{y}) \right] = \hat{\boldsymbol{f}}_A \cdot g_A
\end{aligned}$$

where we replace $\boldsymbol{z} = \boldsymbol{x} \odot \boldsymbol{y} \Leftrightarrow \boldsymbol{z} \odot \boldsymbol{y} = \boldsymbol{x}$. Moreover we use that $w_A(\boldsymbol{z} \odot \boldsymbol{y}) = w_A(\boldsymbol{z}) \cdot w_A(\boldsymbol{y})$. $\square$

A useful insight is that the Rademacher projection can also be obtained by convoluting f with the so-called *Rademacher function* $g_{Rad}(\boldsymbol{x}) := \sum_{i=1}^n x_i$ — the reason is that the Rademacher function serves as a "filter" that precisely keeps the linear part and removes the rest.

Lemma 6.5. *Define $g_{Rad} : \{-1,1\}^n \to \mathbb{R}$ by $g_{Rad}(\boldsymbol{x}) := \sum_{i=1}^n x_i$. Then for any function $f : \{-1,1\}^n \to \mathbb{R}^m$ one has $Rad_n f = f * g_{Rad}$.*

Proof. It suffices to check that the functions on both sides of the equation have identical Fourier coefficients. And in fact, using Lemma 6.4 we see that for $A \subseteq [n]$ one has

$$\widehat{(\mathrm{Rad}_n f)}_A = \begin{cases} \hat{\boldsymbol{f}}_A & \text{if } |A| = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad (\widehat{f * g_{\mathrm{Rad}}})_A = \hat{f}_A \cdot \widehat{g_{\mathrm{Rad}}}_A = \hat{\boldsymbol{f}}_A \cdot \begin{cases} 1 & \text{if } |A| = 1 \\ 0 & \text{otherwise} \end{cases}$$

In fact, both expressions coincide. $\square$

We are still aiming to understand norms of the form $\|f\|_{\mu_n \to \|\cdot\|_K^2}$. At least for the Euclidean norm (i.e. $K = B_2^m$) we have an exact answer. The reason is clear: one can consider $\|f\|_{\mu_n \to \|\cdot\|_2^2}$ as a Euclidean norm on a 2^n-dimensional space. Then the length is invariant if we instead use the coordinates for a different orthonormal basis — and the Walsh functions do form such a basis.

Lemma 6.6. *A function $f : \{-1,1\}^n \to \mathbb{R}^m$ has*

$$\|f\|_{\mu_n \to \|\cdot\|_2^2} = \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} \left[\|f(\boldsymbol{x})\|_2^2\right]^{1/2} \stackrel{!}{=} \Big(\sum_{A \subseteq [n]} \|\hat{\boldsymbol{f}}_A\|_2^2\Big)^{1/2}$$

Proof. We can write the squared norm as

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} \left[\|f(\boldsymbol{x})\|_2^2\right] &= \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} \Big[\Big\| \sum_{A \subseteq [n]} w_A(\boldsymbol{x}) \hat{\boldsymbol{f}}_A \Big\|_2^2\Big] \\ &\stackrel{\|\boldsymbol{z}\|_2^2 = \langle \boldsymbol{z}, \boldsymbol{z}\rangle}{=} \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} \Big[\Big\langle \sum_{A \subseteq [n]} w_A(\boldsymbol{x}) \hat{\boldsymbol{f}}_A, \sum_{B \subseteq [n]} w_B(\boldsymbol{x}) \hat{\boldsymbol{f}}_B \Big\rangle\Big] \\ &= \sum_{A \subseteq [n]} \sum_{B \subseteq [n]} \langle \hat{\boldsymbol{f}}_A, \hat{\boldsymbol{f}}_B \rangle \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n} [w_A(\boldsymbol{x}) \cdot w_B(\boldsymbol{x})]}_{=0 \text{ if } A \neq B, =1 \text{ o.w.}} = \sum_{A \subseteq [n]} \|\hat{\boldsymbol{f}}_A\|_2^2 \end{aligned}$$

$\square$

For a function $g : \{-1,1\}^n \to \mathbb{R}$, we will also use $\|g\|_{\mu_n \to |\cdot|} := \mathbb{E}_{\boldsymbol{x} \in \{-1,1\}^n}[|g(\boldsymbol{x})|]$ as the expected absolute value. We have seen that $\mathrm{Rad}_n f = f * g_{\mathrm{Rad}}$ and so it will be crucial to bound the $\|\cdot\|_{\mu_n \to \|\cdot\|_K^2}$-norm for convolutions.

Lemma 6.7. *Let $K \subseteq \mathbb{R}^m$ be a symmetric convex body and let $f : \{-1,1\}^n \to \mathbb{R}^m$ and $g : \{-1,1\}^n \to \mathbb{R}$ be functions with $g(\boldsymbol{x}) = \sum_{A \subseteq [n]} w_A(\boldsymbol{x}) \cdot c_A$ and $c_A \in \mathbb{R}$. Then the following bounds hold:*

(a) $\|f * g\|_{\mu_n \to \|\cdot\|_K^2} \leq \|f\|_{\mu_n \to \|\cdot\|_K^2} \cdot \|g\|_{\mu_n \to |\cdot|}$.

(b) $\|f * g\|_{\mu_n \to \|\cdot\|_2^2} \leq \|f\|_{\mu_n \to \|\cdot\|_2^2} \cdot \|\boldsymbol{c}\|_\infty$.

(c) $\|f * g\|_{\mu_n \to \|\cdot\|_K^2} \leq \|f\|_{\mu_n \to \|\cdot\|_K^2} \cdot \|\boldsymbol{c}\|_\infty \cdot d_{BM}(K, B_2^m)$.

Proof. For (a) we bound

$$\begin{aligned}
& \|f * g\|_{\mu_n \to \|\cdot\|_K^2} \\
\stackrel{\text{Def}*}{=} & \Big(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n} \Big[\Big\| \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} [f(\boldsymbol{x}\odot\boldsymbol{y}) \cdot g(\boldsymbol{y})] \Big\|_K^2 \Big] \Big)^{1/2} \\
\stackrel{\text{triangle ineq. for } \|\cdot\|_K}{\leq} & \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n} \Big[\Big(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} [\|f(\boldsymbol{x}\odot\boldsymbol{y})\|_K \cdot |g(\boldsymbol{y})|] \Big)^2 \Big]^{1/2} \\
\stackrel{(*)}{\leq} & \Big(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n} \Big[\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} \Big[\|f(\boldsymbol{x}\odot\boldsymbol{y})\|_K^2 \sqrt{|g(\boldsymbol{y})|}^2 \Big] \cdot \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} \Big[\sqrt{|g(\boldsymbol{y})|}^2 \Big] \Big] \Big)^{1/2} \\
= & \Big(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} \Big[|g(\boldsymbol{y})| \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n} \Big[\|f(\boldsymbol{x}\odot\boldsymbol{y})\|_K^2 \Big] \Big] \Big)^{1/2} \cdot \Big(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} [|g(\boldsymbol{y})|] \Big)^{1/2} \\
\stackrel{(**)}{=} & \Big(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} \Big[|g(\boldsymbol{y})| \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n} \Big[\|f(\boldsymbol{x})\|_K^2 \Big] \Big] \Big)^{1/2} \cdot \Big(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^n} [|g(\boldsymbol{y})|] \Big)^{1/2} \\
= & \|f\|_{\mu_n \to \|\cdot\|_K^2} \cdot \|g\|_{\mu_n \to |\cdot|}
\end{aligned}$$

In $(*)$ we use Cauchy-Schwarz in the form $\mathbb{E}[XY] \leq \mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}$ for the random variables $X := \|f(\boldsymbol{x}\odot\boldsymbol{y})\|_K \sqrt{|g(\boldsymbol{y})|}$ and $Y := \sqrt{|g(\boldsymbol{y})|}$ (considering $\boldsymbol{x}$ as fixed; note that one could also use Theorem 1.18 with $\lambda := \frac{1}{2}$). Then in $(**)$ we use that for any fixed $\boldsymbol{y} \in \{-1,1\}^n$ the distribution $\boldsymbol{x}\odot\boldsymbol{y}$ is still uniform from $\{-1,1\}^n$.

For (b), recall that the Fourier expansion of the convolution is of the form $(f * g)(\boldsymbol{x}) = \sum_{A\subseteq[n]} w_A(\boldsymbol{x}) \cdot c_A \cdot \hat{\boldsymbol{f}}_A$ (see Lemma 6.4), hence

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n} \Big[\big\| (f * g)(\boldsymbol{x}) \big\|_2^2 \Big] = \Big(\sum_{A\subseteq[n]} \|c_A \cdot \hat{\boldsymbol{f}}_A\|_2^2 \Big)^{1/2} \leq \|\boldsymbol{c}\|_\infty \cdot \underbrace{\Big(\sum_{A\subseteq[n]} \|\hat{\boldsymbol{f}}_A\|_2^2 \Big)^{1/2}}_{=\mathbb{E}_{\boldsymbol{x}\in\{-1,1\}^n}[\|f(\boldsymbol{x})\|_2^2]^{1/2}}$$

For (c), take a linear map $T : \mathbb{R}^m \to \mathbb{R}^m$ so that $\|\boldsymbol{v}\|_K \leq \|T(\boldsymbol{v})\|_2 \leq d(K, B_2^m) \cdot \|\boldsymbol{v}\|_K$

for every vector $\boldsymbol{v}$ (we will use this twice in $(***)$). Then

$$\begin{aligned}
\|f * g\|_{\mu_n \to \|\cdot\|_K^2} &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\|(f*g)(\boldsymbol{x})\|_K^2\right]^{1/2} \\
&\overset{(***)}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\|T((f*g)(\boldsymbol{x}))\|_2^2\right]^{1/2} \\
&\overset{*\text{ is linear}}{=} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\|(T(f)*g)(\boldsymbol{x})\|_2^2\right]^{1/2} \\
&\overset{(b)+\text{Lem 6.6}}{\leq} \|\boldsymbol{c}\|_\infty \cdot \left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\|T(f(\boldsymbol{x}))\|_2^2\right]\right)^{1/2} \\
&\overset{(***)}{\leq} d_{BM}(K, B_2^m) \cdot \|\boldsymbol{c}\|_\infty \cdot \left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\|f(\boldsymbol{x})\|_K^2\right]\right)^{1/2}
\end{aligned}$$

$\square$

Overall, we already have enough machinery to show that $\kappa(K) \leq d(K, B_2^m)$. For this sake, take a function $f : \{-1,1\}^n \to \mathbb{R}^m$ and use again $g_{\text{Rad}}(\boldsymbol{x}) := \sum_{i=1}^n x_i = \sum_{A\subseteq[n]} w_A(\boldsymbol{x}) c_A$ for $c_A \in \{0,1\}$. Then

$$\|\text{Rad}_n f\|_{\mu_n \to \|\cdot\|_K^2} = \|f * g_{\text{Rad}}\|_{\mu_n \to \|\cdot\|_K^2} \overset{\text{Lem 6.7}}{\leq} \underbrace{\max_{A\subseteq[n]}\{|c_A|\}}_{=1} \cdot d_{BM}(K, B_2^m) \cdot \|f\|_{\mu_n \to \|\cdot\|_K^2}$$

Surprisingly one can reduce that factor to $O(\log d_{BM}(K, B_2^m))$ using more clever arguments.

6.1.2 Constructing almost linear function g with bounded norm

The overall strategy behind Pisier's proof is to find a function $g : \{-1,1\}^n \to \mathbb{R}$ where the linear part is $g_{\text{Rad}}(\boldsymbol{x}) = x_1 + \ldots + x_n$ and then use the triangle inequality to get

$$\|\text{Rad}_n f\|_{\mu_n \to \|\cdot\|_K^2} = \|f * g_{\text{Rad}}\|_{\mu_n \to \|\cdot\|_K^2} \leq \|f * g\|_{\mu_n \to \|\cdot\|_K^2} + \|f * (g - g_{\text{Rad}})\|_{\mu_n \to \|\cdot\|_K^2}$$

In order to bound the first part we need that $\|g\|_{\mu_n \to |\cdot|}$ is small and in order to bound the 2nd part we need g to be almost linear with the linear part being $x_1 + \ldots + x_n$. In fact, such a function exists:

Theorem 6.8 (Pisier). *For any $n, \ell \in \mathbb{N}$ with ℓ odd there is a function $g : \{-1,1\}^n \to \mathbb{R}$ so that one has:*

(I) Bounded norm. *One has $\|g\|_{\mu_n \to |\cdot|} \leq 8\ell$.*

(II) Almost linear. *For $A \subseteq [n]$ one has*

$$\hat{g}_A = \begin{cases} 0 & \text{if } A = \emptyset \\ 1 & \text{if } |A| = 1 \\ 0 & \text{if } 2 \leq |A| \leq \ell \end{cases}$$

and $|\hat{g}_A| \leq \frac{8\ell}{2^\ell}$ for all $|A| > \ell$.

To get some intuition behind the construction, let us first understand why we do not directly choose the linear function $g(\boldsymbol{x}) := x_1 + \ldots + x_n$. The issue is that $\|g\|_{\mu_n \to |\cdot|} = \mathbb{E}_{\boldsymbol{x}\sim\{-1,1\}^n}[|x_1 + \ldots + x_n] = \Theta(\sqrt{n})$ which is far too large. The next idea is to set $g(\boldsymbol{x}) := \prod_{j=1}^n (1 + \varepsilon x_j)$ for some parameter $0 \leq \varepsilon \leq 1$. In this case the norm is small enough as $\|g\|_{\mu_n \to |\cdot|} = \prod_{j=1}^n \mathbb{E}_{x_j \sim \{-1,1\}}[1 + \varepsilon x_j] = 1$. But on the other hand the Fourier coefficients are $\hat{g}_A = \varepsilon^{|A|}$ which is not quite what we need, in particular as $\hat{g}_\emptyset = 1$ while we need a 0. However it turns out that a weighted sum over functions $\prod_{j=1}^n (1 + \varepsilon x_j)$ with different values of ε works.

The first step will be to find a one-dimensional weight function that has the properties that we need. At this point we deviate from the presentation of [AAGM15] and instead use the more elementary and explicit construction of [IRR+20].

Lemma 6.9. *Let $n, \ell \in \mathbb{N}$ with ℓ odd. Then there is a finite set X and functions $\phi : X \to \mathbb{R}$ and $h : X \to [-\frac{1}{2}, \frac{1}{2}]$ so that*

$$\mathbb{E}_{\theta \sim X}\left[\phi(\theta) \cdot h(\theta)^k\right] = \begin{cases} 1 & \text{if } k = 1 \\ 0 & \text{if } k \in \{0, 2, 3, \ldots, \ell\} \end{cases}$$

and $\mathbb{E}_{\theta \sim X}[|\phi(\theta)|] \leq 8\ell$.

Proof. We define $\phi(\theta) := \frac{4\ell - 2}{\ell} \cdot \frac{\sin(\ell\theta)}{\sin^2(\theta)}$, $h(\theta) := \frac{1}{2}\sin(\theta)$, $\Gamma := \{k \cdot \frac{2\pi}{4\ell} \mid k = 0, \ldots, 4\ell - 1\}$ and $X := \Gamma \setminus \{0, \pi\}$. We claim that

$$\mathbb{E}_{\theta \sim \Gamma \setminus \{0,\pi\}}\left[\phi(\theta) \cdot \sin^k(\theta)\right] = \begin{cases} 2 & \text{if } k = 1 \\ 0 & \text{if } k \in \{0, 2, 3, \ldots, \ell\} \end{cases}$$

and $\mathbb{E}_{\theta \sim \Gamma \setminus \{0,\pi\}}[|\phi(\theta)|] \leq 8\ell$ which will satisfy the claim.

We will use 3 simple facts:

- **Fact 1.** One has $\sin(x) = \mathrm{Im}(e^{ix}) = \frac{e^{ix} - e^{-ix}}{2i}$ for $x \in \mathbb{R}$.
- **Fact 2.** For $a \in \mathbb{Z}$ one has $\sum_{\theta \in \Gamma} e^{ia\theta} = \begin{cases} 4\ell & \text{if } a \equiv_{4\ell} 0 \\ 0 & \text{if } a \not\equiv_{4\ell} 0 \end{cases}$

- **Fact 3.** One has $\sum_{j=0}^{\ell-1} z^j = \frac{z^\ell - 1}{z-1}$ for $z \in \mathbb{C} \setminus \{1\}$.

First for $k = 0$ we have $\phi(-\theta) = -\phi(\theta)$ and so $\mathbb{E}_\theta[\phi(\theta)] = 0$ by symmetry. Next, consider $2 \leq k \leq \ell$. In that case

$$\begin{aligned}
\mathop{\mathbb{E}}_{\theta\sim\Gamma\setminus\{0,\pi\}}\left[\phi(\theta)\cdot\sin^k(\theta)\right] &= \frac{1}{|\Gamma|-2}\cdot\frac{4\ell-2}{\ell}\sum_{\theta\in\Gamma\setminus\{0,\pi\}}\sin(\ell\theta)\cdot\sin^{k-2}(\theta) \\
&\overset{\sin(0)=0=\sin(\ell\pi)}{=} \frac{1}{\ell}\sum_{\theta\in\Gamma}\sin(\ell\theta)\cdot\sin^{k-2}(\theta) \\
&\overset{\text{Fact 1}}{=} \frac{1}{\ell}\sum_{\theta\in\Gamma}\Big(\frac{e^{i\ell\theta}-e^{-i\ell\theta}}{2i}\Big)\cdot\Big(\frac{e^{i\ell}-e^{-i\theta}}{2i}\Big)^{k-2} \overset{\text{Fact 2}}{=} 0
\end{aligned}$$

In the last step we use that multiplying out will result in a sum of terms of the form $\sum_{\theta\in\Gamma} e^{ia\theta}$ where $a \in \mathbb{Z}$ with $a \not\equiv_{4\ell} 0$.

Next, we consider the case $k = 1$. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_{\theta\sim\Gamma\setminus\{0,\pi\}}[\phi(\theta)\cdot\sin(\theta)] &= \frac{1}{\ell}\sum_{\theta\in\Gamma\setminus\{0,\pi\}}\frac{\sin(\ell\theta)}{\sin(\theta)} \overset{\text{Fact 1}}{=} \frac{1}{\ell}\sum_{\theta\in\Gamma\setminus\{0,\pi\}}\frac{e^{i\ell\theta}-e^{-i\ell\theta}}{e^{i\theta}-e^{-i\theta}} \\
&\overset{\text{Fact 3}}{=} \frac{1}{\ell}\sum_{\theta\in\Gamma\setminus\{0,\pi\}} e^{-i(\ell-1)\theta}\cdot\frac{(e^{-i2\theta})^\ell-1}{e^{i2\theta}-1} \\
&= \frac{1}{\ell}\sum_{\theta\in\Gamma\setminus\{0,\pi\}}\underbrace{e^{-i(\ell-1)\theta}\sum_{j=0}^{\ell-1}e^{-i2\theta j}}_{=\ell\text{ for }\theta\in\{0,\pi\}} \\
&= \frac{1}{\ell}\Big(-2\ell+\sum_{\theta\in\Gamma}e^{-i(\ell-1)\theta}\sum_{j=0}^{\ell-1}e^{-i2\theta j}\Big) \\
&= \frac{1}{\ell}\Big(-2\ell+\sum_{j=0}^{\ell-1}\underbrace{\sum_{\theta\in\Gamma}e^{-i\theta(\ell-1+2j)}}_{=4\ell\text{ if }\ell-1+2j\equiv_{4\ell}0,=0\text{ o.w.}}\Big) = \frac{1}{\ell}(-2\ell+4\ell) = 2
\end{aligned}$$

where we use that $\ell - 1$ is even and so there is exactly one j so that $\ell - 1 + 2j \equiv_{4\ell} 0$. Finally we bound the average size of the coefficients by

$$\begin{aligned}
\mathop{\mathbb{E}}_{\theta\sim\Gamma\setminus\{0,\pi\}}[|\phi(\theta)|] &\leq \frac{4\ell-2}{\ell}\mathop{\mathbb{E}}_{\theta\sim\Gamma\setminus\{0,\pi\}}\Big[\frac{1}{\sin^2(\theta)}\Big] \\
&\overset{\sin(\theta)\geq\frac{2}{\pi}\theta\ \forall 0\leq\theta\leq\frac{\pi}{2}}{\leq} \frac{4\ell-2}{\ell}\mathop{\mathbb{E}}_{\theta\sim\Gamma\setminus\{0,\pi\}}\Big[\frac{1}{(\frac{2}{\pi}\cdot\theta)^2}\Big] \leq \frac{4}{\ell}\sum_{j=1}^{\ell-1}\frac{1}{(\frac{2}{\pi}\frac{2\pi j}{4\ell})^2} \leq 8\ell
\end{aligned}$$

Where we use symmetry of Γ and sum 4-times over $j = 1, \ldots, \ell - 1$. $\square$

Now we can construct the near-linear function $g : \{-1,1\}^n \to \mathbb{R}$:

Proof of Theorem 6.8. Let X be the set and let ϕ and h be the functions from Lemma 6.9. We choose the function $g : \{-1,1\}^n \to \mathbb{R}$ with

$$g(\boldsymbol{x}) := \mathop{\mathbb{E}}_{\theta\sim X}\left[\phi(\theta)\cdot\prod_{j=1}^{n}\left(1+h(\theta)\cdot x_j\right)\right]$$

Note we can easily read the Fourier coefficients $\hat{g}_A$ as they are simply the coefficients in front of the monomials $\prod_{j\in A} x_j$ when multiplying out. In fact, for any $A \subseteq [n]$ we have

$$\hat{g}_A = \mathop{\mathbb{E}}_{\theta\sim X}\left[\phi(\theta)\cdot h(\theta)^{|A|}\right] = \begin{cases} 0 & \text{if } |A| = 0 \\ 1 & \text{if } |A| = 1 \\ 0 & \text{if } 2 \le |A| \le \ell \end{cases}$$

as required. For $|A| > \ell$ we bound

$$|\hat{g}_A| = \mathop{\mathbb{E}}_{\theta\sim X}\left[|\phi(\theta)|\cdot\underbrace{|h(\theta)|^{|A|}}_{\le 2^{-|A|}}\right] \le \frac{8\ell}{2^{\ell}}.$$

Finally in order to bound the average absolute value we use that always $1 + h(\theta)\cdot x_j > 0$ and so

$$\begin{aligned}
\|g\|_{\mu_n\to|\cdot|} &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\left|\mathop{\mathbb{E}}_{\theta\sim X}\left[\phi(\theta)\cdot\prod_{j=1}^{n}\left(1+h(\theta)\cdot x_j\right)\right]\right|\right] \\
&\le \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\left[\mathop{\mathbb{E}}_{\theta\sim X}\left[|\phi(\theta)|\cdot\prod_{j=1}^{n}\left(1+h(\theta)\cdot x_j\right)\right]\right] \\
&\overset{\text{indep}}{=} \mathop{\mathbb{E}}_{\theta\sim X}\left[|\phi(\theta)|\cdot\prod_{j=1}^{n}\underbrace{\mathop{\mathbb{E}}_{x_j\sim\{-1,1\}}\left[1+h(\theta)\cdot x_j\right]}_{=1}\right] \le \mathop{\mathbb{E}}_{\theta\sim X}[|\phi(\theta)|] \le 8\ell.
\end{aligned}$$

□

6.1.3 Proof of Pisier's Inequality

Finally we come to the main result we have been working towards:

Theorem 6.10 (Pisier)**.** *There is a constant $C > 0$ so that for every symmetric convex body $K \subseteq \mathbb{R}^m$ and $f : \{-1,1\}^n \to \mathbb{R}^m$ one has*

$$\|Rad_n f\|_{\mu_n\to\|\cdot\|_K^2} \le C\log_2(d_{BM}(K, B_2^m)+1)\cdot\|f\|_{\mu_n\to\|\cdot\|_K^2}$$

Proof. Let $\ell \in \mathbb{N}$ be a parameter that we determine later. Let $g = g_{\text{Rad}} + g_{\text{rest}}$ be the approximation of the Rademacher function from the previous Lemma 6.8 where $g_{\text{Rad}}(\boldsymbol{x}) = x_1 + \ldots + x_n$ is the linear part and $g_{\text{rest}}(\boldsymbol{x}) = \sum_{A \subseteq [n]} c_A w_A(\boldsymbol{x})$ with $c_A = 0$ for $|A| \leq \ell$ and $|c_A| \leq \frac{8\ell}{2^\ell}$ for $|A| > \ell$. Now fix any function $f : \{-1,1\}^n \to \mathbb{R}^m$. Then

$$\begin{aligned}
& \|\text{Rad}_n f\|_{\mu_n \to \|\cdot\|_K^2} \\
\stackrel{\text{Lem 6.5}}{=} \quad & \|f * g_{\text{Rad}}\|_{\mu_n \to \|\cdot\|_K^2} \\
\stackrel{\text{triangle ineq}}{\leq} \quad & \|f * g\|_{\mu_n \to \|\cdot\|_K^2} + \|f * g_{\text{rest}}\|_{\mu_n \to \|\cdot\|_K^2} \\
\stackrel{\text{Lem 6.7}}{\leq} \quad & \|f\|_{\mu_n \to \|\cdot\|_K^2} \cdot \underbrace{\|g\|_{\mu_n \to |\cdot|}}_{\leq 8\ell} + d_{BM}(K, B_2^m) \cdot \underbrace{\|\boldsymbol{c}\|_\infty}_{\leq 8\ell/2^\ell} \cdot \|f\|_{\mu_n \to \|\cdot\|_K^2} \\
\leq \quad & \|f\|_{\mu_n \to \|\cdot\|_K^2} \cdot 8\ell \cdot \left(1 + \frac{d_{BM}(K, B_2^m)}{2^\ell}\right)
\end{aligned}$$

then picking ℓ as the smallest odd integer with $\ell \geq \log_2(d(K, B_2^m) + 1)$ gives the claim. $\square$

Phrased differently we have proven that $\kappa(K) \leq O(\log_2(d_{BM}(K, B_2^m) + 1)) \leq O(\log(m))$ for any symmetric convex body $K \subseteq \mathbb{R}^m$.

6.2 Trace duality

Let $\alpha : \mathbb{R}^{n \times n} \to \mathbb{R}_{\geq 0}$ be a *matrix norm*, meaning that α is a norm on the set of $n \times n$ matrices. In particular one has (i) $\alpha(\boldsymbol{V}) = 0 \Leftrightarrow \boldsymbol{V} = \boldsymbol{0}$; (ii) $\alpha(s\boldsymbol{V}) = |s| \cdot \alpha(\boldsymbol{V})$ and (iii) $\alpha(\boldsymbol{V} + \boldsymbol{U}) \leq \alpha(\boldsymbol{V}) + \alpha(\boldsymbol{U})$ where $\boldsymbol{V}, \boldsymbol{U} \in \mathbb{R}^{n \times n}$ and $s \in \mathbb{R}$. Examples would be the *Frobenius norm* $\|\boldsymbol{V}\|_F = (\sum_{i=1}^n \sum_{j=1}^n V_{ij}^2)^{1/2}$ or the *Schatten-p norm* $\|\boldsymbol{V}\|_{S(p)} = (\sum_{i=1}^n \sigma_i(\boldsymbol{V})^p)^{1/p}$ where $\sigma_i(\boldsymbol{V})$ gives the *ith singular value* of $\boldsymbol{V}$ and $p \geq 1$. But back to an arbitrary such matrix norm α. We can always define a *dual norm* by setting

$$\alpha^*(\boldsymbol{V}) := \sup\left\{\text{Tr}[\boldsymbol{V}\boldsymbol{U}] \mid \boldsymbol{U} \in \mathbb{R}^{n \times n} \text{ with } \alpha(\boldsymbol{U}) \leq 1\right\}.$$

It turns out that for any matrix norm there is a matrix so that the norm and the dual norm of the inverse are nicely bounded:

Lemma 6.11 (Lewis [Lew79]). *For any matrix norm $\alpha : \mathbb{R}^{n \times n} \to \mathbb{R}_{\geq 0}$, there exists an invertible matrix $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ so that $\alpha(\boldsymbol{U}) = 1$ and $\alpha^*(\boldsymbol{U}^{-1}) = n$.*

Proof. We can split the proof into two parts.
Claim I. *For any invertible matrix $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ one has $\alpha(\boldsymbol{V}) \cdot \alpha^*(\boldsymbol{V}^{-1}) \geq n$.*

Proof of Claim I. We can write

$$\frac{n}{\alpha(\boldsymbol{V})} = \frac{1}{\alpha(\boldsymbol{V})}\mathrm{Tr}[\boldsymbol{I}_n] = \mathrm{Tr}\Big[\boldsymbol{V}^{-1}\frac{\boldsymbol{V}}{\alpha(\boldsymbol{V})}\Big] \overset{\text{Def dual norm}}{\leq} \alpha^*(\boldsymbol{V}^{-1})$$

which can be then be rearranged to $\alpha(\boldsymbol{V})\cdot\alpha^*(\boldsymbol{V}^{-1}) \geq n$. □

In order to find a matrix that attains equality we will use a *variational argument*. We choose

$$\boldsymbol{U} := \mathrm{argmax}\big\{\det(\boldsymbol{V}) : \boldsymbol{V}\in\mathbb{R}^{n\times n} \text{ invertible with } \alpha(\boldsymbol{V})\leq 1\big\}$$

Claim II. *One has $\alpha(\boldsymbol{U}) = 1$ and $\alpha^*(\boldsymbol{U}^{-1}) \leq n$.*

Proof of Claim II. Clearly we have $\alpha(\boldsymbol{U}) = 1$ and $\det(\boldsymbol{U}) > 0$ as we could scale or flip signs otherwise. It remains to be proven that for any matrix $\boldsymbol{V}\in\mathbb{R}^{n\times n}$ with $\alpha(\boldsymbol{V})\leq 1$, one has $\mathrm{Tr}[\boldsymbol{U}^{-1}\boldsymbol{V}]\leq n$.

So, fix an arbitrary $\boldsymbol{V}\in\mathbb{R}^{n\times n}$ with $\alpha(\boldsymbol{V})\leq 1$. Then for a small enough $\varepsilon > 0$ we have

$$\begin{aligned}
\det\big(\boldsymbol{I}_n + \varepsilon\boldsymbol{U}^{-1}\boldsymbol{V}\big)^{1/n} &= \det\big(\boldsymbol{U}^{-1}\cdot(\boldsymbol{U}+\varepsilon\boldsymbol{V})\big)^{1/n} \qquad (*)\\
&\overset{\text{mult of det}}{=} \alpha(\boldsymbol{U}+\varepsilon\boldsymbol{V})\cdot\underbrace{\det(\boldsymbol{U}^{-1})^{1/n}\cdot\underbrace{\det\Big(\frac{\boldsymbol{U}+\varepsilon\boldsymbol{V}}{\alpha(\boldsymbol{U}+\varepsilon\boldsymbol{V})}\Big)^{1/n}}_{\leq\det(\boldsymbol{U})^{1/n}\text{ by optimality}}}_{\leq 1}\\
&\overset{\alpha\text{ is norm}}{\leq} \underbrace{\alpha(\boldsymbol{U})}_{=1} + \varepsilon\underbrace{\alpha(\boldsymbol{V})}_{\leq 1} = 1+\varepsilon
\end{aligned}$$

Considering the derivative of the determinant function[1] we see that

$$\varepsilon \overset{(*)}{\geq} \det(\boldsymbol{I}_n + \varepsilon\boldsymbol{U}^{-1}\boldsymbol{V})^{1/n} - 1 \overset{\text{up to } O(\varepsilon^2)\text{ terms}}{\approx} \varepsilon\frac{\mathrm{Tr}[\boldsymbol{U}^{-1}\boldsymbol{V}]}{n}$$

Then for $\varepsilon\to 0$ we obtain $\mathrm{Tr}[\boldsymbol{U}^{-1}\boldsymbol{V}]\leq n$ as needed. □

6.3 The ℓ-norm

Let $K\subseteq\mathbb{R}^n$ be a symmetric convex body. We define the *ℓ-norm* of a matrix $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ as

$$\ell_K(\boldsymbol{A}) := \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\Big[\|\boldsymbol{A}\boldsymbol{x}\|_K^2\Big]^{1/2}$$

(first introduced by [FTJ79]). We can verify that ℓ_K is indeed a norm.

[1] Consider $f(t) := \det(\boldsymbol{I}_n + t\boldsymbol{M})$ with $\boldsymbol{M}\in\mathbb{R}^{n\times n}$. Then $f'(0) = \mathrm{Tr}[\boldsymbol{M}]$.

Lemma 6.12. *ℓ_K is a norm.*

Proof. We only verify the triangle inequality, meaning that for matrices $\boldsymbol{A},\boldsymbol{B} \in \mathbb{R}^{n\times n}$, we need to prove that $\ell_K(\boldsymbol{A}+\boldsymbol{B}) \leq \ell_K(\boldsymbol{A}) + \ell_K(\boldsymbol{B})$. Then

$$\begin{aligned}\ell_K(\boldsymbol{A}+\boldsymbol{B}) &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\left\|(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}\right\|_K^2\right]^{1/2} \overset{\|\cdot\|_K \text{ norm}}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\left(\|\boldsymbol{A}\boldsymbol{x}\|_K + \|\boldsymbol{B}\boldsymbol{x}\|_K\right)^2\right]^{1/2} \\ &\overset{(*)}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\|\boldsymbol{A}\boldsymbol{x}\|_K^2\right]^{1/2} + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\|\boldsymbol{B}\boldsymbol{x}\|_K^2\right]^{1/2} = \ell_K(\boldsymbol{A}) + \ell_K(\boldsymbol{B}).\end{aligned}$$

In $(*)$, we use *Minkowski's inequality* from Lemma 1.20 for $p=2$ and the jointly distributed random variables $X := \|\boldsymbol{A}\boldsymbol{x}\|_K$ and $Y := \|\boldsymbol{B}\boldsymbol{x}\|_K$ where $\boldsymbol{x}\sim\gamma_n$. ☐

To give some relation to known quantities, later we will argue that $\ell_K(\boldsymbol{I}_n) = \Theta(\sqrt{n})\cdot M(K)$. Recall that $\boldsymbol{A} \succ 0$ means that $\boldsymbol{A}$ is symmetric and all Eigenvalues are strictly positive. Then finding a matrix $\boldsymbol{A}$ with $\boldsymbol{A}\succ 0$ so that $\ell_K(\boldsymbol{A})\cdot\ell_{K^\circ}(\boldsymbol{A}^{-1})$ is small is essentially the same as finding a linear transformation $T:\mathbb{R}^n\to\mathbb{R}^n$ so that $M(T(K))\cdot M(T(K)^\circ)$ is small. We can make this relation formal:

Lemma 6.13. *Let $K\subseteq\mathbb{R}^n$ be a symmetric convex body and let $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ with $\boldsymbol{A}\succ 0$. Then (i) $\ell_K(\boldsymbol{A}) = \ell_{A^{-1}(K)}(\boldsymbol{I}_n)$, (ii) $\ell_{K^\circ}(\boldsymbol{A}^{-1}) = \ell_{(A^{-1}(K))^\circ}(\boldsymbol{I}_n)$ and (iii) $\ell_K^*(\boldsymbol{A}^{-1}) = \ell^*_{A^{-1}(K)}(\boldsymbol{I}_n)$.*

Proof. We have $\|\boldsymbol{A}\boldsymbol{x}\|_K = \|\boldsymbol{x}\|_{A^{-1}(K)}$ which gives (i). Next, $\|\boldsymbol{A}^{-1}\boldsymbol{x}\|_{K^\circ} = \|\boldsymbol{x}\|_{A(K^\circ)} = \|\boldsymbol{x}\|_{(A^{-1}(K))^\circ}$ proving (ii). For (iii) we have $\sup\{\mathrm{Tr}[\boldsymbol{A}^{-1}\boldsymbol{U}] : \ell_K(\boldsymbol{U})\leq 1\} = \sup\{\mathrm{Tr}[\boldsymbol{I}_n\boldsymbol{V}] : \ell_K(\boldsymbol{A}\boldsymbol{V})\leq 1\} = \sup\{\mathrm{Tr}[\boldsymbol{I}_n\boldsymbol{V}] : \ell_{A^{-1}(K)}(\boldsymbol{V})\leq 1\}$ where we make the substitution $\boldsymbol{V} := \boldsymbol{A}^{-1}\boldsymbol{U} \Leftrightarrow \boldsymbol{U} = \boldsymbol{A}\boldsymbol{V}$. ☐

The next step is to find a good candidate for the matrix $\boldsymbol{A}$ that has a bounded value of $\ell_K(\boldsymbol{A})\cdot\ell_{K^\circ}(\boldsymbol{A}^{-1})$.

Lemma 6.14 (Consequence of Lewis Lemma)**.** *For any symmetric convex body $K\subseteq\mathbb{R}^n$ there is a matrix $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ with $\boldsymbol{A}\succ 0$ so that $\ell_K(\boldsymbol{A}) = 1$ and $\ell_K^*(\boldsymbol{A}^{-1}) = n$.*

Proof. As ℓ_K is a matrix norm, Lewis Lemma (Lemma 6.11) guarantees us a matrix $\boldsymbol{B}\in\mathbb{R}^{n\times n}$ with $\ell_K(\boldsymbol{B}) = 1$ and $\ell_K^*(\boldsymbol{B}^{-1}) = n$. But that matrix $\boldsymbol{B}$ does not have to be symmetric. So we consider the *singular value decomposition* $\boldsymbol{B} = \sum_{k=1}^n \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$ where $\boldsymbol{u}_1,\dots,\boldsymbol{u}_n$ and $\boldsymbol{v}_1,\dots,\boldsymbol{v}_n$ are two orthonormal bases of $\mathbb{R}^n$. We set $\boldsymbol{A} := \sum_{k=1}^n \sigma_k \boldsymbol{u}_k \boldsymbol{u}_k^T$ which by construction is symmetric and positive definite. We note that for $\boldsymbol{x}\sim\gamma_n$, the random vectors $\boldsymbol{B}\boldsymbol{x} = \sum_{k=1}^n \sigma_k \boldsymbol{u}_k \langle \boldsymbol{v}_k, \boldsymbol{x}\rangle$ and $\boldsymbol{A}\boldsymbol{x} = \sum_{k=1}^n \sigma_k \boldsymbol{u}_k \langle \boldsymbol{u}_k, \boldsymbol{x}\rangle$ have identical distributions. Hence $\ell_K(\boldsymbol{A}) = \ell_K(\boldsymbol{B})$. Also reinspecting the proof Lemma 6.11 we recall that $\boldsymbol{B}$ was chosen to maximize $\det(\boldsymbol{B})$, subject to $\ell_K(\boldsymbol{B})\leq 1$. But $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$ and so also $\boldsymbol{A}$ is a suitable choice. ☐

The reader may have noted the crucial difference that the quantity $\ell_K(\boldsymbol{A})$ is defined via a random Gaussian while the framework around the version of Pisier's inequality that we have presented, uses random variables from $\{-1,1\}$. But we can use the *Central Limit Theorem* to approximate a Gaussian arbitrarily well with sums of $\{-1,1\}$-random variables. For a large enough N, construct a matrix $\boldsymbol{B} \in \mathbb{R}^{nN\times n}$, which for any $i \in [n]$ contains N times the column vector $\frac{1}{\sqrt{N}}\boldsymbol{e}_i$. Then we know that $\boldsymbol{B}\boldsymbol{y}$ with $\boldsymbol{y} \sim \{-1,1\}^{nN}$ is a good approximation to the standard Gaussian distribution γ_n. To be more concrete, we have a continuous function $F: \mathbb{R}^n \to \mathbb{R}$ and we need that $\mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[F(\boldsymbol{x})] = \lim_{N\to\infty}\mathbb{E}_{\boldsymbol{y}\in\{-1,1\}^{nN}}[F(\boldsymbol{B}\boldsymbol{y})]$. However, the continuity is not enough. For example if $n=1$ and $F(x) = e^{x^3}$, then $\mathbb{E}_{x\sim\gamma_1}[F(x)] = \infty$ while $\mathbb{E}_{\boldsymbol{y}\in\{-1,1\}^N}[F(\boldsymbol{B}\boldsymbol{y})]$ is finite (though growing in N). However, it suffices to add as additional condition that that function increases moderately. The version of the Central Limit Theorem that suffices for our purpose is the following:

Theorem 6.15 (Implication of the Central Limit Theorem)**.** *Let $F: \mathbb{R}^n \to \mathbb{R}$ be a continuous function with $\lim_{\|\boldsymbol{x}\|_1\to\infty} F(\boldsymbol{x})\cdot e^{-\|\boldsymbol{x}\|_1} = 0$. Fix a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and let $\boldsymbol{B} \in \mathbb{R}^{nN\times n}$ be the matrix that contains N column vectors of the form $\frac{\boldsymbol{A}^i}{\sqrt{N}}$ for each $i \in [n]$. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[F(\boldsymbol{A}\boldsymbol{x})] = \lim_{N\to\infty}\mathop{\mathbb{E}}_{\boldsymbol{y}\in\{-1,1\}^{nN}}\left[F(\boldsymbol{B}\boldsymbol{y})\right]$$

Now we come to handle the main challenge: we know how to obtain a matrix $\boldsymbol{A}$ with $\boldsymbol{A} \succ 0$ so that $\ell_K(\boldsymbol{A}) = 1$ and $\ell_K^*(\boldsymbol{A}^{-1}) = n$. But what we really need is an upper bound on $\ell_{K^\circ}(\boldsymbol{A}^{-1})$. This is the next step and this is the part where we crucially rely on Pisier's inequality.

Lemma 6.16. *For any convex symmetric body $K \subseteq \mathbb{R}^n$ and any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $\boldsymbol{A} \succ 0$ one has $\ell_{K^\circ}(\boldsymbol{A}^{-1}) \le \kappa(K)\cdot\ell_K^*(\boldsymbol{A}^{-1})$.*

Proof. Following Lemma 6.13 we can replace K by $\tilde{K} := A^{-1}(K)$ and it suffices to prove the claim $\ell_{K^\circ}(\boldsymbol{I}_n) \le \kappa(K)\cdot\ell_K^*(\boldsymbol{I}_n)$. For a large enough N, we make the same choice of $\boldsymbol{B} \in \mathbb{R}^{nN\times n}$ which contains N copies of $\frac{\boldsymbol{e}_i}{\sqrt{N}}$ as columns. We define a function $f: \{-1,1\}^{nN} \to \mathbb{R}^n$ with $f(\boldsymbol{y}) := \boldsymbol{B}\boldsymbol{y}$. Then we have

$$\ell_{K^\circ}(\boldsymbol{I}_n) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\|\boldsymbol{x}\|_{K^\circ}^2\right]^{1/2} \overset{(*)}{\approx} \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\{-1,1\}^{nN}}\Big[\|\underbrace{\boldsymbol{B}\boldsymbol{y}}_{=:f(\boldsymbol{y})}\|_{K^\circ}^2\Big]^{1/2} = \|f\|_{\mu_{nN}\to\|\cdot\|_{K^\circ}^2}$$

by the Central Limit Theorem (Theorem 6.15), where we can make the multiplicative or additive error in $(*)$ as small as we like by choosing N large enough.

Next, we introduce a function $h : \{-1,1\}^{nN} \to \mathbb{R}^n$ where for $\boldsymbol{y} \in \{-1,1\}^{nN}$, we choose $h(\boldsymbol{y})$ as an element that is a multiple of the *dual element* to $f(\boldsymbol{y})$ with respect to the norm $\|\cdot\|_{K^\circ}$. More precisely, we choose $h(\boldsymbol{y}) \in \mathbb{R}^m$ so that

$$\|h(\boldsymbol{y})\|_K = \frac{\|f(\boldsymbol{y})\|_{K^\circ}}{\|f\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}}} \quad \text{and} \quad \langle f(\boldsymbol{y}), h(\boldsymbol{y})\rangle = \frac{\|f(\boldsymbol{y})\|^2_{K^\circ}}{\|f\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}}}$$

This choice is always possible, see Lemma 1.7. Then

$$\|h\|_{\mu_{nN}\to\|\cdot\|^2_K} = \mathop{\mathbb{E}}_{\boldsymbol{y}\in\{-1,1\}^{nN}}\left[\|h(\boldsymbol{y})\|^2_K\right]^{1/2} = \frac{\mathbb{E}_{\boldsymbol{y}\in\{-1,1\}^{nN}}\left[\|f(\boldsymbol{y})\|^2_{K^\circ}\right]^{1/2}}{\|f\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}}} = 1.$$

Moreover

$$\langle f,h\rangle_{\mu_{nN}} = \frac{\mathbb{E}_{\boldsymbol{y}\sim\{-1,1\}^{nN}}[\|f(\boldsymbol{y})\|^2_{K^\circ}]}{\|f\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}}} = \|f\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}}$$

In other words, the function h is the dual element to f with respect to the norm $\|\cdot\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}}$. We note that f is a linear function and the non-zero Fourier-coefficients are $\hat{\boldsymbol{f}}_{\{j\}} = \boldsymbol{B}^j$. We want to remark that we do not know whether h is linear. Let $[nN] = J_1 \dot\cup \ldots \dot\cup J_n$ be the partition of indices so that $\hat{\boldsymbol{f}}_{\{j\}} = \boldsymbol{B}^j = \frac{\boldsymbol{e}_i}{\sqrt{N}}$ for all $j \in J_i$. We note that h is symmetric in the sense that $h(\boldsymbol{y})$ is invariant under permuting indices $j, j' \in J_i$. This implies that $\hat{\boldsymbol{h}}_{\{j\}} = \hat{\boldsymbol{h}}_{\{j'\}}$ for all $j, j' \in J_i$. We define a matrix $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ so that $\hat{\boldsymbol{h}}_{\{j\}} = \frac{\boldsymbol{C}^i}{\sqrt{N}}$ for $j \in I_i$ and $i \in [n]$. Then

$$\begin{aligned}
\ell_{K^\circ}(\boldsymbol{I}_n) \quad &\approx \quad \|f\|_{\mu_{nN}\to\|\cdot\|^2_{K^\circ}} \overset{h \text{ is dual el. of } f}{=} \langle f,h\rangle_{\mu_{nN}} \\
&= \quad \sum_{S\subseteq[nN]} \underbrace{\langle\hat{\boldsymbol{f}}_S, \hat{\boldsymbol{h}}_S\rangle}_{=0 \text{ if } |S|\neq 1 \text{ as } f \text{ linear}} \\
&= \quad \sum_{j=1}^{nN} \langle\hat{\boldsymbol{f}}_{\{j\}}, \hat{\boldsymbol{h}}_{\{j\}}\rangle \\
&\overset{(**)}{=} \quad \sum_{i=1}^{n} \langle\boldsymbol{e_i}, \boldsymbol{C}^i\rangle = \mathrm{Tr}[\boldsymbol{I}_n\boldsymbol{C}] \overset{\text{Cauchy-Schwarz}}{\leq} \ell^*_K(\boldsymbol{I}_n)\cdot\ell_K(\boldsymbol{C})
\end{aligned}$$

where we use in $(**)$ we use that each coordinate i appears N times. Then using the Central Limit Theorem again

$$\ell_K(\boldsymbol{C}) \overset{CLT}{\approx} \mathop{\mathbb{E}}_{\boldsymbol{y}\in\{-1,1\}^{nN}}\left[\left\|\sum_{j=1}^{nN} y_j\hat{\boldsymbol{h}}_{\{j\}}\right\|^2_K\right]^{1/2} = \|\mathrm{Rad}_{nN}(h)\|_{\mu_{nN}\to\|\cdot\|^2_K} \leq \kappa(K)\cdot\underbrace{\|h\|_{\mu_{nN}\to\|\cdot\|^2_K}}_{=1}$$

$\square$

Now we can finish the proof of the $\ell\ell^{\circ}$-estimate:

Theorem 6.17 ($\ell\ell^{\circ}$-Estimate [FTJ79])**.** *For any convex symmetric body $K \subseteq \mathbb{R}^n$ there is a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $\boldsymbol{A} \succ 0$ so that $\ell_K(\boldsymbol{A}) \cdot \ell_{K^{\circ}}(\boldsymbol{A}^{-1}) \leq n \cdot \kappa(K)$.*

Proof. We obtain a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ from Lemma 6.14 with $\boldsymbol{A} \succ 0$ so that $\ell_K(\boldsymbol{A}) = 1$ and $\ell_K^*(\boldsymbol{A}^{-1}) = n$ where ℓ_K^* is the dual norm of ℓ_K. Then applying Lemma 6.16 gives $\ell_{K^{\circ}}(\boldsymbol{A}^{-1}) \leq \kappa(K) \cdot \ell_K^*(\boldsymbol{A}^{-1}) \leq \kappa(K) \cdot n$, which then provides the claim. □

6.4 The MM°-estimate

Recall that the M-value of a symmetric convex body K is

$$M(K) := \mathop{\mathbb{E}}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_K]$$

and the dual estimate is $M^{\circ}(K) := M(K^{\circ})$ (we however prefer using $M(K^{\circ})$ to keep notation at a minimum). More generally one can define $M_p(K) := \mathbb{E}_{\boldsymbol{x}\in S^{n-1}}\left[\|\boldsymbol{x}\|_K^p\right]^{1/p}$; then one can in fact proof that $M_1(K) \leq M_2(K) \leq \sqrt{2} \cdot M_1(K)$, which would imply that

$$M(K) = \Theta\Big(\frac{1}{\sqrt{n}}\Big) \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\|\boldsymbol{x}\|_K^2\right]^{1/2}$$

We prove this in more generality but without determining the exact constant.

Lemma 6.18. *There is a constant $C > 0$ so that for $p \geq 1$ and any symmetric convex body $K \subseteq \mathbb{R}^n$ one has $M_1(K) \leq M_p(K) \leq C\sqrt{p} \cdot M_1(K)$.*

Proof. If we define the random variable $X := \|\boldsymbol{x}\|_K$ with $\boldsymbol{x} \sim S^{n-1}$, then the lower bound is equivalent to $\mathbb{E}[X]^p \leq \mathbb{E}[X^p]$ which holds true by *Jensen's inequality*. For the upper bound, we recall *Kahane's Inequality* from Theorem3.19 which shows that for $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ one has

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^m}\left[\left\|\boldsymbol{A}\boldsymbol{x}\right\|_K^p\right]^{1/p} \leq O(\sqrt{p}) \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}\in\{-1,1\}^m}\left[\left\|\boldsymbol{A}\boldsymbol{x}\right\|_K\right]$$

Making the substitution from Theorem 6.15, the same inequality also holds for Gaussians[2]. □

Theorem 6.19 (MM°-estimate)**.** *For any symmetric convex body $K \subseteq \mathbb{R}^n$, there is an invertible linear map T so that $M(T(K)) \cdot M(T(K)^{\circ}) \leq O(1) \cdot \kappa(K)$.*

[2] Alternatively, the proof of Kahane's inequality could be easily adapted to directly hold for Gaussians.

Proof. We can apply a linear transformation to K so that $\ell_K(\boldsymbol{I}_n)\cdot\ell_{K^\circ}(\boldsymbol{I}_n) \leq n\cdot\kappa(K)$. Then $\ell_K(\boldsymbol{I}_n) = \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_K^2]^{1/2} = \Theta(\sqrt{n})\cdot M(K)$ and $\ell_{K^\circ}(\boldsymbol{I}_n) = \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_{K^\circ}^2]^{1/2} = \Theta(\sqrt{n})\cdot M(K^\circ)$. Together that gives the claim. □

Recall that any symmetric convex body K with $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(B_2^n)$ must have a *mean width* of $w(K) \geq 1$ (see Theorem 1.28). On the other hand, the mean width can be arbitrarily high for a symmetric convex body. But it turns out that there is always a linear transformation that achieves a bound of $O(\log(n))$ as always $\kappa(K) \leq O(\log n)$.

Theorem 6.20 (Reverse Urysohn Inequality)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then there exists a linear map T so that $\mathrm{Vol}_n(T(K)) = \mathrm{Vol}_n(B_2^n)$ and $w(T(K)) \leq O(\kappa(K))$.*

Proof. We first prove a non-trivial claim:

Claim. *Let $Q \subseteq \mathbb{R}^n$ be any convex body with $\mathbf{0} \in int(Q)$. Then $w(K) \geq 2\Big(\frac{Vol_n(B_2^n)}{Vol_n(Q^\circ)}\Big)^{1/n}$.*

Proof of Claim. Recall the quantities $h_Q(\boldsymbol{\theta}) = \max\{\langle\boldsymbol{\theta},\boldsymbol{x}\rangle : \boldsymbol{x} \in Q\}$ and $\rho_Q(\boldsymbol{\theta}) := \max\{r \geq 0 \mid r\boldsymbol{\theta} \in Q\}$ for $\boldsymbol{\theta} \in S^{n-1}$ and the fact that $h_Q(\boldsymbol{\theta}) = \frac{1}{\rho_{Q^\circ}(\boldsymbol{\theta})}$. Then

$$w(Q) = 2\mathop{\mathbb{E}}_{\boldsymbol{\theta}\sim S^{n-1}}[h_Q(\boldsymbol{\theta})] \overset{(*)}{\geq} 2\mathop{\mathbb{E}}_{\boldsymbol{\theta}\sim S^{n-1}}\big[h_Q(\boldsymbol{\theta})^{-n}\big]^{-1/n} = \Big(\mathop{\mathbb{E}}_{\boldsymbol{\theta}\sim S^{n-1}}[\rho_{Q^\circ}(\boldsymbol{\theta})^n]\Big)^{-1/n} \overset{(**)}{=} \Big(\frac{\mathrm{Vol}_n(Q^\circ)}{\mathrm{Vol}_n(B_2^n)}\Big)^{-1/n}$$

Here for the inequality $(*)$ we can verify that for a positive random variable X one has $\mathbb{E}[X] \geq \mathbb{E}[X^{-n}]^{-1/n}$ since by Hölders Inequality (Theorem 1.18) we can bound $(\mathbb{E}[X]\cdot\mathbb{E}[X^{-n}]^{\frac{1}{n}})^{\frac{n}{n+1}} = \mathbb{E}[X]^{\frac{n}{n+1}}\,\mathbb{E}[X^{-n}]^{\frac{1}{n+1}} \geq \mathbb{E}[X^{\frac{n}{n+1}}X^{-n\cdot\frac{1}{n+1}}] = \mathbb{E}[X^0] = 1$. Finally we have used integration in polar coordinates from Lemma 1.46 to obtain $(**)$. □

Now, using the MM°-estimate one can find a linear map T so that $w(T(K))\cdot w(T(K)^\circ) \leq O(\kappa(K))$ (recall that $w(Q) = 2M(Q^\circ)$ for any symmetric convex body). After scaling we may further assume that $\mathrm{Vol}_n(T(K)) = \mathrm{Vol}_n(B_2^n)$. Then

$$w(T(K)^\circ) \overset{\text{Claim}}{\geq} 2\Big(\frac{\mathrm{Vol}_n(B_2^n)}{\mathrm{Vol}_n(T(K))}\Big)^{1/n} = 2$$

from which it follows that $w(T(K)) \leq O(\kappa(K))$. □

We can also provide a result for non-symmetric bodies:

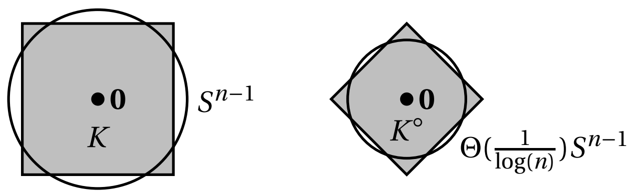
Theorem 6.21. *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then there is a linear transformation T so that $\mathrm{Vol}_n(T(K)) = \mathrm{Vol}_n(B_2^n)$ and $w(T(K)) \leq O(\log(n))$.*

Proof. Since the difference body $K-K$ is symmetric, we can use the last Theorem and apply a linear transformation to K so that $\mathrm{Vol}_n(K-K) = 1$ and $w(K-K) \leq$

$O(\log(n))$. We already noted in an earlier chapter that $w(K-K) = 2w(K)$. Moreover, $\mathrm{Vol}_n(K) \geq 4^{-n} \cdot \mathrm{Vol}_n(K-K)$ by Rogers-Shephard inequality. Then scaling K by some factor in $[1,4]$ will satisfy the claim. $\square$

6.5 A geometric interpretation of the ℓ-position

We want to give a different geometric interpretation of the ℓ-position that feels a bit like a relaxed variant of the John position. We will see that every symmetric convex body can be put into a position where it contains 90% of the sphere S^{n-1} while the polar of K contains 90% of the smaller sphere $\Theta(\frac{1}{\log(n)})S^{n-1}$. The constants are of course chosen arbitrarily.



More formally:

Theorem 6.22. *For any symmetric convex body $K \subseteq \mathbb{R}^n$, there is a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ so that*

$$\Pr_{\boldsymbol{x}\sim S^{n-1}}[\boldsymbol{x} \in T(K)] \geq \frac{9}{10} \quad \textit{and} \quad \Pr_{\boldsymbol{x}\sim \frac{1}{C\log(n)}S^{n-1}}[\boldsymbol{x} \in T(K)^\circ] \geq \frac{9}{10}$$

where $C > 0$ is a universal constant.

Proof. Using the MM°-estimate of Theorem 6.19, we may apply a linear transformation to the body K so that $M(K) = \frac{1}{10}$ and $M(K^\circ) \leq \frac{C}{10}\log(n)$ for some constant $C > 0$. Then $M(K) = \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_K] = \frac{1}{10}$ and by Markov's inequality, $\Pr_{\boldsymbol{x}\sim S^{n-1}}[\boldsymbol{x} \notin K] = \Pr_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_K > 1] \leq \frac{1}{10}$. Similarly $M(K^\circ) = \mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_{K^\circ}] \leq \frac{C}{10}\log(n)$ and again by Markov's inequality, $\Pr_{\boldsymbol{x}\sim S^{n-1}}[\boldsymbol{x} \notin C\log(n)\cdot K^\circ] = \Pr_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_{K^\circ} > C\log(n)] \leq \frac{1}{10}$. $\square$

6.6 Computing the ℓ-position

We want to briefly comment on the issue whether the ℓ-position of a body is computable in polynomial time. Let K be a symmetric convex body. Again, consider the quantity $\ell_K(\boldsymbol{A}) = \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{A}\boldsymbol{x}\|_K^2]^{1/2}$ for a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. Then with the same arguments as in the beginning of Chapter 2 we can replace $\boldsymbol{A}$ with a positive semi-definite matrix $\boldsymbol{B}$ so that $\ell_K(\boldsymbol{A}) = \ell_K(\boldsymbol{B})$. That means, finding the matrix $\boldsymbol{A}$ for Theorem 6.17 boils down to solving the *convex program*

$$\begin{array}{rcl} \max \det(\boldsymbol{A}) & & (CP1) \\ \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}\left[\|\boldsymbol{A}\boldsymbol{x}\|_K^2\right]^{1/2} & \leq & 1 \\ \boldsymbol{A} & \succeq & 0 \end{array}$$

Let $\boldsymbol{A}$ be the optimum solution. Now set $\tilde{K} := A^{-1}(K)$. We consider again the convex program, but this time for $\tilde{K}$:

$$\begin{array}{rcl} \max \det(\boldsymbol{B}) & & (CP2) \\ \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}\left[\|\boldsymbol{B}\boldsymbol{x}\|_{\tilde{K}}^2\right]^{1/2} & \leq & 1 \\ \boldsymbol{B} & \succeq & 0 \end{array}$$

Note that $\|\boldsymbol{B}\boldsymbol{x}\|_{\tilde{K}} = \|\boldsymbol{B}\boldsymbol{x}\|_{A^{-1}(K)} = \|\boldsymbol{B}\boldsymbol{A}\boldsymbol{x}\|_K$. Then it is not hard to see that $\boldsymbol{B} = \boldsymbol{I}_n$ is the optimum solution to $(CP2)$. From the proof of Theorem 6.17 we know that $\ell_{\tilde{K}}(\boldsymbol{I}_n) = 1$ and $\ell_{\tilde{K}^\circ}(\boldsymbol{I}_n) \leq O(n \log d_{BM}(K))$. We can rewrite this as $\mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_{\tilde{K}}^2]^{1/2} = 1$ and $\mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_{\tilde{K}^\circ}] \leq O(n \log d_{BM}(K))$.

Exercises

Exercise 6.1.
Consider the Schatten-p norm $\alpha(\boldsymbol{V}) := \|\boldsymbol{V}\|_{\mathcal{S}(p)}$ for $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ where $p \geq 1$. Find a matrix $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ so that $\alpha(\boldsymbol{U}) = 1$ and $\alpha^*(\boldsymbol{U}^{-1}) = n$.

Exercise 6.2.
For $K := B_\infty^n$, compute (up to universal constants) the quantities $\ell_K(\boldsymbol{I}_n)$, $\ell_K^*(\boldsymbol{I}_n)$ and $\ell_{K^\circ}(\boldsymbol{I}_n)$.

Exercise 6.3.
In the following let $K \subseteq \mathbb{R}^n$ be any convex symmetric body.

(i) Prove that for any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ one has $\ell_{K^\circ}^{\mathrm{Rad}}(\boldsymbol{A}) \leq \kappa(K) \cdot (\ell_K^{\mathrm{Rad}})^*(\boldsymbol{A})$.

(ii) Prove that for any convex symmetric body $K \subseteq \mathbb{R}^n$ there is a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ so that $\ell_K^{\mathrm{Rad}}(\boldsymbol{A}) \cdot \ell_{K^\circ}^{\mathrm{Rad}}(\boldsymbol{A}^{-1}) \leq n \cdot \kappa(K)$.

Exercise 6.4.

Let n be a power of 2. In this exercise we want to prove that $K := B_\infty^n$ has $\kappa(K) \geq \Omega(\sqrt{\log(n)})$ (which by the way is tight for unconditional norms). For a complex number $z = a+bi \in \mathbb{C}$ with $a, b \in \mathbb{R}$ we denote $\Re(z) = a$ and $\Im(z) = b$. Let $d := \log_2(n)$ and consider the function $h : \{-1,1\}^d \to \mathbb{R}$ defined by $h(\boldsymbol{x}) := \Im(\prod_{j=1}^d (1 + \frac{x_j}{\sqrt{d}} \cdot i))$.

(i) Prove that $|h(\boldsymbol{x})| \leq O(1)$ for all $\boldsymbol{x} \in \{-1,1\}^d$.

(ii) Prove that $\hat{h}_{\{j\}} = \frac{1}{\sqrt{d}}$ for all $j \in [d]$.

Now consider $f : \{-1,1\}^d \to \mathbb{R}^n$ where we index coordinates of the vector $f(\boldsymbol{x}) \in \mathbb{R}^n$ by $f(\boldsymbol{x})_{\boldsymbol{y}}$ for $\boldsymbol{y} \in \{-1,1\}^d$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ we denote $\boldsymbol{x} \odot \boldsymbol{y} \in \mathbb{R}^d$ as the coordinate wise product, that means $(\boldsymbol{x} \odot \boldsymbol{y})_j := x_j \cdot y_j$. Then define the function by $f(\boldsymbol{x})_{\boldsymbol{y}} := h(\boldsymbol{x} \odot \boldsymbol{y})$.

(iii) Prove that $\|f(\boldsymbol{x})\|_\infty \leq O(1)$ for all $\boldsymbol{x} \in \{-1,1\}^d$ and $\|f\|_{\mu_d \to \|\cdot\|_\infty^2} \leq O(1)$.

(iv) Prove that $\hat{\boldsymbol{f}}_{\{j\}} = \frac{1}{\sqrt{d}} \cdot (y_j)_{\boldsymbol{y} \in \{-1,1\}^d}$ for all $j \in [d]$.

(v) Prove that for every $\boldsymbol{x} \in \{-1,1\}^d$ one has $\|\mathrm{Rad}_d f(\boldsymbol{x})\|_\infty \geq \sqrt{d}$ and $\|\mathrm{Rad}_d f\|_{\mu_d \to \|\cdot\|_\infty^2} \geq \sqrt{d}$.

(vi) Conclude that $\kappa(B_\infty^n) \geq \Omega(\sqrt{\log(n)})$.

Chapter 7

The Quotient of Subspaces Theorem and Linear Duality

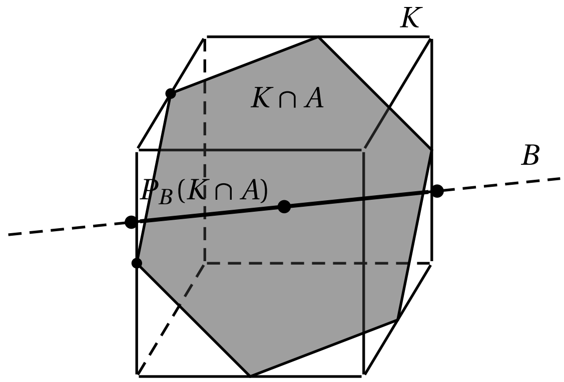
Suppose that $K \subseteq \mathbb{R}^n$ is a symmetric convex body. We denote $\text{radius}(K) := \max\{\|\boldsymbol{x}\|_2 : \boldsymbol{x} \in K\}$. We want to revisit the question of finding $O(1)$-ellipsoidal slices of K. We know already that there might not be any subspace $A \subseteq \mathbb{R}^n$ with $\dim(A) \geq \omega(\log(n))$ so that $K \cap A$ is $O(1)$-ellipsoidal. Also projections are not necessarily better. We have seen in Lemma 5.23 that for a body $K \subseteq B_2^n$, a random k-dimensional subspace A will have $\text{radius}(P_A(K)) \leq O(M(K^\circ) + \sqrt{k/n})$. Here it might be helpful to recall that $M(K^\circ) = \frac{1}{2}w(K)$. Again, if $k \gg \log(n)$ is large, this bound might not be good enough. It turns out that we can prove a stronger result:

> *Pajor-Tomczak Theorem. Let $K \subseteq \mathbb{R}^n$ be a symmetric convex set and let F be a random λn-dimensional subspace. Then $\text{radius}(K \cap F) \leq O(\frac{1}{\sqrt{1-\lambda}}) \cdot M(K^\circ)$ with high probability.*

However, this result only gives a one-sided bound. It was Milman who proved that the *combination* of intersection and projection indeed leads to near ellipsoidal sets. In a simplified form one can prove that:

> *Milman's Quotient of Subspaces Theorem. For any symmetric convex body $K \subseteq \mathbb{R}^n$ there are subspaces $B \subseteq A \subseteq \mathbb{R}^n$ with $\dim(B) \geq n/2$ and an ellipsoid $\mathcal{E} \subseteq B$ so that $\mathcal{E} \subseteq P_B(K \cap A) \subseteq O(1) \cdot \mathcal{E}$.*

Of course this is a high dimensional phenomenom — an attempt at a visualization for $K \subseteq \mathbb{R}^3$ is as follows:



Towards the end of the chapter we will prove another remarkable result that gives a geometric duality-type relation between a body K and its polar:

> *Milman's Linear Duality Theorem. For any symmetric convex body K there is a subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) \geq \Omega(n)$ so that either radius$(K \cap F) \leq O(1)$ or radius$(K^\circ \cap F) \leq O(1)$.*

7.1 A simple bound on the radius of $K \cap A$

First, we want to work towards the Pajor-Tomczak Theorem. The goal is to obtain a bound that radius$(K \cap A) \leq f(\lambda) \cdot M(K^\circ)$ holds with high probability where $A \subseteq \mathbb{R}^n$ is a random subspace with $\dim(A) = \lambda n$. It is clear that for $\lambda \to 1$, one must have $f(\lambda) \to \infty$. Note that later we will indeed have use for the regime where $\lambda \approx 1$. One can also see that the Pajor-Tomczak bound is optimal as for an $(n - \Theta(1))$-dimensional subspace it gives radius$(K \cap A) \leq O(\sqrt{n}) \cdot M(K^\circ)$ which is best possible.

To warm up, we give a simple, yet suboptimal bound.

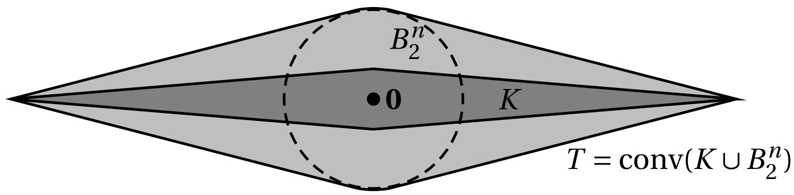
Theorem 7.1. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $0 < \lambda < 1$. Draw a random λn-dimensional subspace $A \subseteq \mathbb{R}^n$. Then*

$$\Pr\left[\textit{radius}(K \cap A) \leq C^{1/(1-\lambda)} \cdot M(K^\circ)\right] \geq 1 - e^{-n}$$

where $C > 0$ is a universal constant.

Proof. After scaling we may assume that $M(K^\circ) = 1$. The idea is to apply the *Volume Ratio Theorem* (Theorem 5.19) which guarantees that one has radius$(K \cap A) \leq (O(\alpha))^{n/(n-\dim(A))}$ with high probability, provided that $B_2^n \subseteq K$ and $\mathrm{Vol}_n(K) = \alpha^n \cdot \mathrm{Vol}_n(B_2^n)$.

The first issue is that K might be arbitrarily thin in some directions and hence not contain B_2^n. We fix this by considering instead the enlarged set $T := \mathrm{conv}(K \cup B_2^n)$ which is again convex and symmetric.



But potentially the larger set T has a larger M°-value and we need to show that $M(T^\circ)$ is not too large. In fact

$$M(T^\circ) = \frac{1}{2}w(T) \overset{\mathrm{conv}(K\cup B_2^n)\subseteq K+B_2^n}{\leq} \frac{1}{2}w(K+B_2^n) = \underbrace{\frac{1}{2}w(K)}_{=M(K^\circ)=1} + \underbrace{\frac{1}{2}w(B_2^n)}_{=1} = 2$$

The next step is to obtain an upper bound on the volume of T. For this purpose we make use of *Urysohn's Inequality* (Theorem 1.28) to get

$$2 \geq M(T^\circ) = \frac{1}{2}w(T) \overset{\text{Urysohn}}{\geq} \left(\frac{\mathrm{Vol}_n(T)}{\mathrm{Vol}_n(B_2^n)}\right)^{1/n}$$

This can be rearranged to $\mathrm{Vol}_n(T) \leq 2^n \cdot \mathrm{Vol}_n(B_2^n)$. Since now $B_2^n \subseteq T$, we can apply the *Volume Ratio Theorem* and conclude that a random λn-dimensional subspace $A \subseteq \mathbb{R}^n$ satisfies $\mathrm{radius}(T \cap A) \leq (c \cdot 2)^{\frac{n}{n-\lambda n}}$ with probability at least $1 - 2^{-n}$. The claim follows from $K \subseteq T$. $\square$

7.2 The Theorem of Pajor-Tomczak

Finally we show the radius result with the optimum dependence. We want to recall Lemma 5.23 that we have proven in Section 5.7.

> *Lemma 5.23. For any centrally symmetric convex set $K \subseteq t \cdot B_2^n$, a random k-dimensional subspace $A \subseteq \mathbb{R}^n$ satisfies*
>
> $$\Pr\left[\mathit{radius}(P_A(K)) \leq O(1) \cdot \max\left\{M(K^\circ), t\sqrt{\tfrac{k}{n}}\right\}\right] \geq 1 - e^{-\Omega(k)}$$

We we state and prove the Theorem of Pajor and Tomczak-Jaegermann (following the later proof of [Mil90a]):

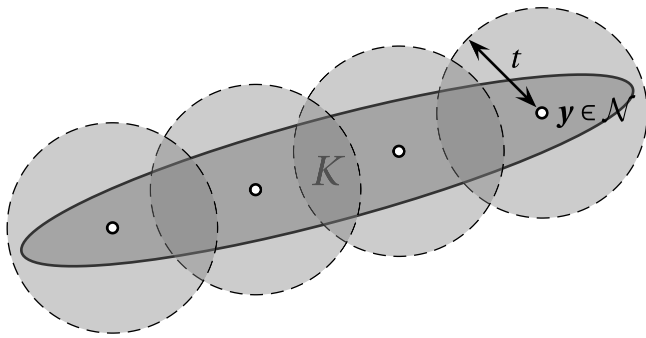
Theorem 7.2 (Pajor, Tomczak-Jaegermann [PTJ86])**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex set and let $0 < \varepsilon < 1$. Then if we draw a uniform random $(1-\varepsilon)n$-dimensional subspace A, then*

$$\Pr\left[\mathit{radius}(K \cap A) \leq O\left(\frac{1}{\sqrt{\varepsilon}}\right) \cdot M(K^\circ)\right] \geq 1 - e^{-\Theta(1)\cdot\varepsilon n}.$$

Proof. For a large enough constant $C > 0$, we pick a parameter $t := C\sqrt{1/\varepsilon} \cdot M(K^\circ)$ and let $\mathcal{N} \subseteq K$ be a minimum cardinality t-net of K, meaning that $K \subseteq \bigcup_{\boldsymbol{y}\in\mathcal{N}}(\boldsymbol{y} + tB_2^n)$. Then *Sudakov's Inequality* (Theorem 4.12) bounds the size of that net as

$$|\mathcal{N}| = N(K, tB_2^n) \leq \exp\left(\Theta(n) \cdot (w(K)/t)^2\right) \leq e^{\varepsilon n/100},$$

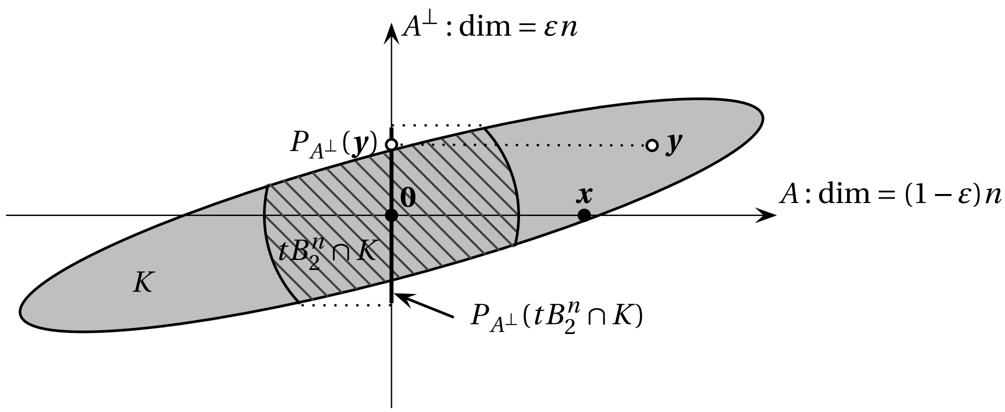
if C is chosen large enough, using that $w(K) = 2M(K^\circ)$.



The trick behind the proof is to study the the projection of K into $A^\perp$ instead of directly considering the intersection $K \cap A$. For a large enough constant C', we consider two events

$$\begin{aligned}
\mathcal{A} &:= \text{“}\|P_{A^\perp}(\boldsymbol{y})\|_2 \geq \frac{1}{2}\sqrt{\varepsilon} \cdot \|\boldsymbol{y}\|_2 \;\; \forall \boldsymbol{y} \in \mathcal{N}\text{”} \\
\mathcal{B} &:= \text{“radius}\big(P_{A^\perp}(K \cap tB_2^n)\big) \leq C' \cdot M(K^\circ)\text{”}
\end{aligned}$$

which basically say that the projection of certain parts of K into $A^\perp$ is close enough to the expected value.



Claim I. *One has* $\Pr[\mathcal{A} \,\&\, \mathcal{B}] \geq 1 - e^{-\Theta(\varepsilon n)}$.
Proof of Claim I. First fix a unit length vector $\boldsymbol{y} \in S^{n-1}$. Then one has $\mathbb{E}[\|P_{A^\perp}(\boldsymbol{y})\|_2^2] =$

$\frac{\dim(A)}{n} = \frac{\varepsilon n}{n} = \varepsilon$. Moreover, the function $\boldsymbol{y} \mapsto \|P_{A^\perp}(\boldsymbol{y})\|_2$ is 1-Lipschitz. Hence by the usual concentration argument $\Pr_A[\|P_{A^\perp}(\boldsymbol{y})\|_2 \geq \frac{1}{2}\sqrt{\varepsilon}] \geq 1 - e^{-\varepsilon n/10}$. As $|\mathcal{N}| \leq e^{\varepsilon n/100}$ we obtain $\Pr[\mathcal{A}] \geq 1 - e^{-\Omega(\varepsilon n)}$ from the union bound. For event $\mathcal{B}$ we invoke Lemma 5.23 to obtain that with probability $1 - e^{-\Theta(\varepsilon n)}$ one has

$$\text{radius}(P_{A^\perp}(K \cap tB_2^n)) \leq O(1) \cdot \max\Big\{M(K^\circ), \underbrace{t\sqrt{\tfrac{\dim(A^\perp)}{n}}}_{=C\cdot M(K^\circ)}\Big\} \leq C' \cdot M(K^\circ)$$

if C' is chosen large enough. The claim follows. $\square$

Claim II. *If events $\mathcal{A}$ & $\mathcal{B}$ happen, then radius$(K \cap A) \leq O(\frac{1}{\sqrt{\varepsilon}}) \cdot M(K^\circ)$.*

Proof of claim II. Fix a point $\boldsymbol{x} \in K \cap A$ and let $\boldsymbol{y} \in \mathcal{N}$ be the net point with $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq t$. Note that $\boldsymbol{x} - \boldsymbol{y} \in 2K$ by the triangle inequality and so $\boldsymbol{x} - \boldsymbol{y} \in tB_2^n \cap 2K \subseteq 2(tB_2^n \cap K)$. Then $\boldsymbol{x} \in A$ means that $P_{A^\perp}(\boldsymbol{x}) = \mathbf{0}$ and so

$$\|P_{A^\perp}(\boldsymbol{y})\|_2 = \|P_{A^\perp}(\boldsymbol{x} - \boldsymbol{y})\|_2 \leq 2 \cdot \text{radius}\big(P_{A^\perp}(tB_2^n \cap K)\big) \stackrel{\text{event } \mathcal{B}}{\leq} 2C' \cdot M(K^\circ)$$

Then the length of $\boldsymbol{x}$ can be bounded as

$$\|\boldsymbol{x}\|_2 \stackrel{\text{triangle ineq.}}{\leq} \|\boldsymbol{y}\|_2 + \underbrace{\|\boldsymbol{x} - \boldsymbol{y}\|_2}_{\leq t} \stackrel{\text{event } \mathcal{A}}{\leq} \frac{2}{\sqrt{\varepsilon}}\|P_{A^\perp}(\boldsymbol{y})\|_2 + t \leq O\Big(\frac{1}{\sqrt{\varepsilon}}\Big) \cdot M(K^\circ)$$

as claimed. $\square$

7.3 The Quotient of Subspaces Theorem

Recall that for a symmetric convex set K and the ball B_2^n, the *Banach-Mazur distance* $d_{BM}(K, B_2^n)$ is the minimum number $R \geq 1$ so that there is an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ so that $\mathcal{E} \subseteq K \subseteq R \cdot \mathcal{E}$. Note that always $d_{BM}(K, B_2^n) \leq \sqrt{n}$. An important ingredient of the Quotient of Subspaces Theorem is the following "One-step argument". In particular it implies already that there are subspaces $\mathbb{R}^n \supseteq A \supseteq B$ with $\dim(B) \geq \frac{n}{2}$ where $\mathcal{E} \subseteq P_B(K \cap A) \subseteq O(\log n) \cdot \mathcal{E}$ for some ellipsoid $\mathcal{E}$.

We also want to refresh a couple of facts on polarity. If $K \subseteq Q$ are convex bodies containing $\mathbf{0}$ in the interior, then $K^\circ \supseteq Q^\circ$. For any $R > 0$, the polar of the radius-R ball is $(R \cdot B_2^n)^\circ = \frac{1}{R} B_2^n$. Moreover, if $K \subseteq \mathbb{R}^n$ is a symmetric convex body and $H \subseteq \mathbb{R}^n$ is a subspace, then $(K \cap H)^\circ = P_H(K^\circ)$ (see Lemma 1.9). Note that here for the first time in this text we are using polarity for convex sets that are not full-dimensional.

Theorem 7.3 (One-step Quotient Subspace)**.** *For every symmetric convex set $K \subseteq \mathbb{R}^n$ and $0 < \lambda < 1$, there are subspaces $\mathbb{R}^n \supseteq A \supseteq B$ with $\dim(B) \geq \lambda^2 n$ and an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ so that $\mathcal{E} \subseteq P_B(K \cap A) \subseteq O(\frac{1}{1-\lambda}) \cdot \ln(d_{BM}(K, B_2^n) + 1) \cdot \mathcal{E}$.*

Proof. The claim is invariant under linear transformations, hence we may assume that the body K is in *ℓ-position*, which according to Theorem 6.19 means that $M(K)\cdot M(K^\circ) \leq O(\ln(d_{BM}(K,B_2^n)+1)$. We sample a λn-dimensional subspace $A \subseteq \mathbb{R}^n$ and from Pajor-Tomczak (Theorem 7.2) we know that radius$(K\cap A) \leq O(\frac{1}{\sqrt{1-\lambda}})M(K^\circ)$ with probability $1-e^{-\Theta(\lambda n)}$. This can be equivalently written as

$$K\cap A \subseteq \Theta\Big(\frac{M(K^\circ)}{\sqrt{1-\lambda}}\Big)\cdot(B_2^n\cap A) \qquad (I)$$

Taking the polar on both sides of the inclusion in (I) gives

$$(K\cap A)^\circ \supseteq \Theta\Big(\frac{\sqrt{1-\lambda}}{M(K^\circ)}\Big)\cdot(B_2^n\cap A) \qquad (II)$$

We take again a $\lambda^2 n$-dimensional random subspace $B\subseteq A$ and again with probability $1-e^{-\Theta(\lambda^2 n)}$ the claim from the Pajor-Tomczak Theorem applies and

$$\text{radius}((K\cap A)^\circ\cap B) \leq O\Big(\frac{M(((K\cap A)^\circ)^\circ)}{\sqrt{1-\lambda}}\Big) = O\Big(\frac{M(K\cap A)}{\sqrt{1-\lambda}}\Big) \leq O\Big(\frac{M(K)}{\sqrt{1-\lambda}}\Big) \qquad (III)$$

Here we have used that $((K\cap A)^\circ)^\circ = K\cap A$ and by Theorem 5.15 also $M(K\cap A)\leq O(1)\cdot M(K)$ with high probability over the choice of A. Then

$$\Theta\Big(\frac{\sqrt{1-\lambda}}{M(K^\circ)}\Big)\cdot(B_2^n\cap B) \overset{(II)}{\subseteq} (K\cap A)^\circ\cap B \overset{(III)}{\subseteq} O\Big(\frac{M(K)}{\sqrt{1-\lambda}}\Big)\cdot(B_2^n\cap B) \qquad (IV)$$

Taking the polars of (IV)[1] we obtain

$$\Theta\Big(\frac{\sqrt{1-\lambda}}{M(K)}\Big)\cdot(B_2^n\cap B) \subseteq P_B(K\cap A) \subseteq \Theta\Big(\frac{M(K^\circ)}{\sqrt{1-\lambda}}\Big)\cdot(B_2^n\cap B)$$

That means $P_B(K\cap A)$ is near-spherical with approximation factor $O(\frac{1}{1-\lambda})\cdot M(K^\circ)\cdot M(K) \leq O(\frac{1}{1-\lambda})\cdot\ln(d_{BM}(K,B_2^n)+1)$ using that K is in ℓ-position. The claim then follows. $\square$

Theorem 7.3 already provides the surprising claim that there are subspaces $\mathbb{R}^n \supseteq A \supseteq B$ with $\dim(B)\geq n/2$ so that $d_{BM}(P_B(K\cap A), B_2^{\dim(B)}) \leq O(\log n)$. However, we can improve the logarithmic factor to $O(1)$ by *iterating* the argument. In particular suppose we have a sequence $\mathbb{R}^n \supseteq A_0 \supseteq B_0 \supseteq A_1 \supseteq B_1 \supseteq \ldots \supseteq A_{T-1} \supseteq B_{T-1}$ of nested subspaces and consider the iterate $K_{t+1} := P_{B_t}(K_t\cap A_t)$ for $t\in\{0,\ldots,T-1\}$. Then there are also subspaces $\mathbb{R}^n\supseteq A^*\supseteq B^*$ with $\dim(B^*)\geq\dim(B_{T-1})$ so that the final iterate K_T can be obtained by a single intersection/projection, i.e. $K_T = P_{B^*}(K\cap A^*)$. This follows from the following lemma:

[1] Note that taking the polars of $K_1\subseteq K_2\subseteq K_3$ means obtaining the relation $K_3^\circ\subseteq K_2^\circ\subseteq K_1^\circ$; also recall that we are using Lemma 1.9 here.

Lemma 7.4. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then*

(a) For subspaces $\mathbb{R}^n \supseteq A \supseteq B$ one has $P_B(K \cap A) = P_{B+A^\perp}(K) \cap B$.

(b) For subspaces $\mathbb{R}^n \supseteq A \supseteq B \supseteq C \supseteq D$ one has $P_D(P_B(K \cap A) \cap C) = P_D(K \cap A \cap B^\perp)$.

Proof. For (a), we can check that

$$\begin{aligned} P_{B+A^\perp}(K) \cap B &= \{\boldsymbol{x} \in B \mid \boldsymbol{x}+\boldsymbol{y} \in K \text{ with } \boldsymbol{y} \perp B \text{ and } \boldsymbol{y} \perp A^\perp\} \\ &= \{\boldsymbol{x} \in B \mid \boldsymbol{x}+\boldsymbol{y} \in K \text{ with } \boldsymbol{y} \in A \cap B^\perp\} \\ &\overset{\boldsymbol{x}+\boldsymbol{y}\in A \text{ anyway}}{=} \{\boldsymbol{x} \in B \mid \boldsymbol{x}+\boldsymbol{y} \in K \cap A \text{ with } \boldsymbol{y} \in B^\perp\} = P_B(K \cap A) \end{aligned}$$

For (b), we can see that

$$\begin{aligned} P_D(P_B(K \cap A) \cap C) &\overset{B \supseteq C}{=} P_D(P_{C+(B^\perp)^\perp}(K \cap A) \cap C) \overset{(a)}{=} P_D(P_C((K \cap A) \cap B^\perp)) \\ &\overset{C \supseteq D}{=} P_D(K \cap A \cap B^\perp) \end{aligned}$$

□

The full version of Milman's Theorem [Mil85] requires Theorem 7.3 plus 1-2 pages of calculations to get the tight estimates. Hence we prefer a weaker but simpler-to-prove and more explicit bound that is sufficient for our later applications:

Theorem 7.5 (Simple Quotient of Subspaces Theorem)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then there are subspaces $\mathbb{R}^n \supseteq A \supseteq B$ with $\dim(B) \geq \frac{n}{2}$ so that $d_{BM}(P_B(K \cap A), B_2^{\dim(B)}) \leq O(1)$.*

Proof. We will call Theorem 7.3 iteratively. We set $K_0 := K$ and $D_0 := d_{BM}(K, B_2^n)$. In each iteration $t \in \{1, 2, \ldots\}$ we will maintain a body $K_t \subseteq B_{t-1}$ that is an iterated intersection/projection of the original body K and B_{t-1} is some subspace. As the original body is full-dimensional we will have $\dim(K_t) = \dim(B_{t-1})$. We abbreviate $D_t := d_{BM}(K_t, B_2^{\dim(K_t)})$ as the Banach-Mazur distance of the intermediate body. If that distance satisfies $D_t \leq 100C^3$ then we set $T := t$ and terminate the procedure with K_T as final iterate; here $C \geq 1$ is the unspecified constant from Theorem 7.3. Otherwise, if D_t is above that threshold, then we apply the procedure from Theorem 7.3 to find subspaces $A_t \supseteq B_t$ so that we can set $K_{t+1} := P_{B_t}(K_t \cap A_t)$. For any parameter $0 < \delta_t < 1$ that we choose in a moment, we have the guarantee that $D_{t+1} \leq \frac{C}{\sqrt{\delta_t}} \cdot \ln(10D_t)$ and $\dim(K_{t+1}) \geq \dim(K_t) \cdot (1-\delta_t)$. The choice for parameters that we make is $\delta_t := \frac{1}{D_t}$ so that indeed

$$D_{t+1} \leq C\sqrt{D_t} \cdot \ln(10D_t) \overset{D_t \geq 100C^3}{\leq} \frac{1}{2} D_t$$

as D_t was large enough. Note that $\delta_{T-1} = \frac{1}{D_{T-1}} \leq \frac{1}{100C^3}$. Also observe that the δ_t-values are geometrically increasing; more precisely we have $\delta_{T-1} \geq 2\delta_{T-2} \geq 4\delta_{T-3} \geq \ldots \geq 2^{T-1}\delta_0$. The dimension of the body at the end of the procedure is

$$\begin{aligned}
\dim(K_T) &= n\prod_{t=0}^{T-1}(1-\delta_t) \geq n \cdot \exp\Big(-2\sum_{t=0}^{T-1}\delta_t\Big) \geq n \cdot \exp\Big(-2\delta_{T-1}\underbrace{\sum_{i=0}^{T-1} 2^i}_{\leq 2}\Big) \\
&\geq n \cdot \exp(-4\delta_{T-1}) \overset{\delta_{T-1} \leq \frac{1}{100}}{\geq} n \cdot \exp(-25)
\end{aligned}$$

The claim follows. $\square$

We have so far avoided an explanation, why the Theorem of Milman is called *Quotient of Subspaces Theorem.* Consider the normed vector space $X = (\mathbb{R}^n, \|\cdot\|_K)$ and take subspaces $\mathbb{R}^n \supseteq A \supseteq B$. We can consider the *quotient subspace* $A \setminus B := A \cap B^{\perp}$ and equip it with the norm $\|\cdot\|_{P_B(K \cap A)}$. Then Milman's Theorem says that there is a quotient subspace of high dimension that has a bounded Banach-Mazur distance to the corresponding ℓ_2-space. However, we have prefered a geometric presentation of the topic.

7.4 The Bourgain-Milman Inequality

In Chapter 1 we have discussed the *Blaschke-Santaló inequality* which proves that among all symmetric convex bodies $K \subseteq \mathbb{R}^n$, the Euclidean ball maximizes the *Mahler product*

$$s(K) := \mathrm{Vol}_n(K) \cdot \mathrm{Vol}_n(K^{\circ})$$

In other words, one always has $s(K) \leq s(B_2^n)$. It is somewhat surprising that this is basically an equality in the sense that there is a small universal constant $C > 0$ so that for every symmetric convex body $K \subseteq \mathbb{R}^n$ one has

$$C^n \leq \frac{\mathrm{Vol}_n(K) \cdot \mathrm{Vol}_n(K^{\circ})}{\mathrm{Vol}_n(B_2^n)^2} \leq 1.$$

That means $\text{Vol}_n(K)$ determines $\text{Vol}_n(K^\circ)$ up to a factor of $2^{\Theta(n)}$. It might be worth mentioning that the Mahler product is invariant under linear transformations, which means that for any bijective linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ one has $s(A(K)) = s(K)$. We now show this result due to Bourgain and Milman which is a nice application of the Quotient of Subspaces Theorem that we have just proven (we will see a second proof in Chapter 8 using M-ellipsoids). Here we will use repetedly $(K \cap F)^\circ = P_F(K^\circ)$ from Lemma 1.9.

Theorem 7.6 (Bourgain-Milman [BM87]). *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex body. Then*

$$\frac{\mathit{Vol}_n(K) \cdot \mathit{Vol}_n(K^\circ)}{\mathit{Vol}_n(B_2^n)^2} \geq 2^{-\Theta(n)}$$

We proof the statement by an iterative / inductive argument over the quantity

$$\alpha_N := \inf\Big\{\Big(\frac{s(K)}{s(B_2^n)}\Big)^{1/n} \mid K \subseteq \mathbb{R}^n \text{ is symmetric convex body with } n \leq N\}$$

Then the goal is prove that $\alpha_N \geq \Omega(1)$. For the sake of simplicity suppose we have a body K that attains this value for some n, i.e. $\alpha_n = (s(K)/s(B_2^n))^{1/n}$. We use the simple version of the *Quotient of Subspace Theorem* from Theorem 7.5 to obtain two subspaces $\mathbb{R}^n \supseteq A \supseteq B$ with $\dim(B) = \frac{n}{2}$ and $\dim(A) =: m$ so that $P_B(K \cap A)$ is within a factor C_0 of an ellipsoid. As the claim is invariant under linear transformations, so we may assume that

$$(B_2^n \cap B) \subseteq P_B(K \cap A) \subseteq C_0 \cdot (B_2^n \cap B)$$

In wise foresight, we define lower dimensional bodies

$$K_1 := P_{A^\perp}(K) \quad \text{and} \quad K_2 := K \cap A \cap B^\perp.$$

Note that by Lemma 1.9,

$$K_1^\circ = (P_{A^\perp}(K))^\circ = K^\circ \cap A^\perp \quad \text{and} \quad K_2^\circ = (K \cap (A \cap B^\perp))^\circ = P_{A \cap B^\perp}(K^\circ) = P_{B^\perp}(P_A(K^\circ))$$

are the polar of K_1 with respect to subspace $A^\perp$ and the polar of K_2 with respect to subspace $A \cap B^\perp$.

The key ingredient for the proof is the insight from Lemma 1.48 that the volume of a symmetric convex body can be well approximated by taking the product of the volume of an intersection and a projection into the complementary subspace. We use this insight twice to estimate

$$\begin{aligned}
& \mathrm{Vol}_n(K) & (*) \\
\overset{\text{Lem 1.48 for } K \text{ and subsp. } A}{\geq} \; & 2^{-2n} \cdot \mathrm{Vol}_m(K \cap A) \cdot \mathrm{Vol}_{n-m}(P_{A^\perp}(K)) \\
\overset{\text{Lem 1.48 for } K\cap A \text{ \& s.p. } B^\perp}{\geq} \; & 2^{-4n} \mathrm{Vol}_{n/2}(\underbrace{P_B(K \cap A)}_{\supseteq B_2^n \cap B}) \cdot \mathrm{Vol}_{m-n/2}(\underbrace{K \cap A \cap B^\perp}_{=K_2}) \cdot \mathrm{Vol}_{n-m}(\underbrace{P_{A^\perp}(K)}_{K_1}) \\
\geq \; & 2^{-4n} \cdot \mathrm{Vol}_{n/2}(B_2^{n/2}) \cdot \mathrm{Vol}_{m-n/2}(K_2) \cdot \mathrm{Vol}_{n-m}(K_1)
\end{aligned}$$

We apply the same line of arguments to the polar of K:

$$\begin{aligned}
& \mathrm{Vol}_n(K^\circ) & (**) \\
\overset{\text{Lem 1.48 for } K \text{ and subsp. } A^\perp}{\geq} \; & 2^{-2n} \cdot \mathrm{Vol}_{n-m}(K^\circ \cap A^\perp) \cdot \mathrm{Vol}_m(P_A(K^\circ)) \\
\overset{\text{Lem 1.48 for } P_A(K^\circ) \text{ \& s.p. } B}{\geq} \; & 2^{-4n} \mathrm{Vol}_{n-m}(\underbrace{K^\circ \cap A^\perp}_{=K_1^\circ}) \cdot \mathrm{Vol}_{n/2}(\underbrace{P_A(K^\circ) \cap B}_{=(P_B(K\cap A))^\circ}) \cdot \mathrm{Vol}_{m-n/2}(\underbrace{P_{B^\perp}(P_A(K^\circ))}_{=K_2^\circ}) \\
\geq \; & 2^{-4n} \mathrm{Vol}_{n-m}(K_1^\circ) \cdot C_0^{-n/2} \mathrm{Vol}_{n/2}(B_2^{n/2}) \cdot \mathrm{Vol}_{m-n/2}(K_2^\circ)
\end{aligned}$$

using that $(P_B(K \cap A))^\circ = (K \cap A)^\circ \cap B = P_A(K^\circ) \cap B$ by applying Lemma 1.9 twice. Then multiplying the bounds in $(*)$ and $(**)$ and abbreviating the unit ball volume $\kappa_n := \mathrm{Vol}_n(B_2^n)$ we obtain

$$\begin{aligned}
\alpha_n^n \quad = \quad & \frac{\mathrm{Vol}_n(K) \cdot \mathrm{Vol}_n(K^\circ)}{\mathrm{Vol}_n(B_2^n)^2} & (***) \\
\overset{(*)+(**)}{\geq} \quad & 2^{-8n} C_0^{-n/2} \frac{\kappa_{n/2}^2}{\kappa_n^2} \cdot \underbrace{\mathrm{Vol}_{n-m}(K_1)\mathrm{Vol}_{n-m}(K_1^\circ)}_{\geq \alpha_n^{m-n} \cdot \kappa_{n-m}^2} \cdot \underbrace{\mathrm{Vol}_{m-n/2}(K_2)\mathrm{Vol}_{m-n/2}(K_2^\circ)}_{\geq \alpha_n^{m-n/2} \cdot \kappa_{m-n/2}^2} \\
\geq \quad & 2^{-8n} C_0^{-n/2} \Big(\underbrace{\frac{\kappa_{n/2} \kappa_{m-n/2} \kappa_{n-m}}{\kappa_n}}_{\geq 1} \Big)^2 \cdot \alpha_n^{n-m} \alpha_n^{m-n/2} \\
\geq \quad & 2^{-8n} \cdot C_0^{-n/2} \cdot \alpha_n^{n/2}
\end{aligned}$$

Here we use that α_n must be monotonically non-decreasing. Moreover, we use that satisfies $\kappa_{r+s} \leq \kappa_r \cdot \kappa_s$ for all $r, s > 0$ as one can easily see from the upper bound of Lemma 1.48. Then rearranging $(***)$ for α_n gives that $\alpha_n \geq 2^{-16} C_0^{-1}$ which then gives the claim.

This brings us to a beautiful duality theorem for covering numbers:

Theorem 7.7 (Duality of Covering Numbers — König-Milman [KM87])**.** *For every pair $K, T \subseteq \mathbb{R}^n$ of symmetric, convex bodies one has*

$$2^{-\Theta(n)} N(T^\circ, K^\circ) \leq N(K, T) \leq 2^{\Theta(n)} N(T^\circ, K^\circ)$$

Proof. Again we use that for symmetric convex sets S, Q one has $N(S,Q) = 2^{\Theta(n)} \cdot \frac{\mathrm{Vol}_n(S+\frac{1}{2}Q)}{\mathrm{Vol}_n(\frac{1}{2}Q)}$. Then we get the relation between covering numbers and volume ratios:

$$\begin{aligned} N(T^\circ, K^\circ) \quad &\leq \quad 2^{\Theta(n)} \frac{\mathrm{Vol}_n(T^\circ + K^\circ)}{\mathrm{Vol}_n(K^\circ)} \overset{(*)}{\leq} 2^{\Theta(n)} \frac{\mathrm{Vol}_n((K \cap T)^\circ)}{\mathrm{Vol}_n(K^\circ)} \\ &\overset{(**)}{=} \quad 2^{\Theta(n)} \frac{\mathrm{Vol}_n(K)}{\mathrm{Vol}_n(K \cap T)} \overset{(***)}{\leq} 2^{\Theta(n)} N(K, T). \end{aligned}$$

In $(*)$ we use that $T^\circ + K^\circ \subseteq 2\mathrm{conv}(K^\circ \cup T^\circ) = 2((T^\circ)^\circ \cap (K^\circ)^\circ)^\circ = 2(T \cap K)^\circ$ by Lemma 1.8. In $(**)$ we use the inequalities of Bourgain-Milman (Theorem 7.6) and Blaschke-Santaló (Theorem 1.30). In $(***)$ we use that $\mathrm{Vol}_n(K \cap T) \cdot N(K,T) \geq \mathrm{Vol}_n(K)$ because for symmetric convex bodies the intersection $\mathrm{Vol}_n((\boldsymbol{x} + T) \cap K)$ is maximized for $\boldsymbol{x} = \mathbf{0}$. $\square$

7.5 The Linear Duality Relation

We have already proven earlier in Theorem 5.11 that for every symmetric convex body K, a random subspace F of dimension $\dim(F) \geq \Omega(\sqrt{n})$ satisfies the following: either $K \cap F$ or $K^\circ \cap F$ is $O(1)$-spherical. We will prove here that we can improve the dimension of the subspace F to $\Omega(n)$ if we are satisfied with a *one-sided* bound that guarantees either $K \cap F$ or $K^\circ \cap F$ to be contained in a ball of radius $O(1)$.

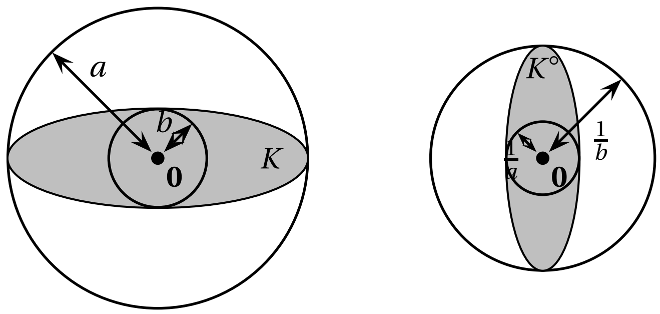
The Distance Lemma

We begin by proving a seemingly unrelated lemma that is called the *Distance Lemma*. We know that for K symmetric and convex and $\boldsymbol{x} \in S^{n-1}$ we have $\|\boldsymbol{x}\|_K \cdot \|\boldsymbol{x}\|_{K^\circ} \geq \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 1$ by Cauchy-Schwarz. The Distance Lemma on the other hand will give us a upper bound on $\|\boldsymbol{x}\|_K$ and $\|\boldsymbol{x}\|_{K^\circ}$, depending on the radius and inradius of K. To get some intuition we first state a simplified version:

Lemma 7.8 (Simplified Distance Lemma)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex symmetric set with $\frac{1}{r} B_2^n \subseteq K \subseteq r B_2^n$. Then $\|\boldsymbol{x}\|_K^2 + \|\boldsymbol{x}\|_{K^\circ}^2 \leq r^2 + 1$ for all $\boldsymbol{x} \in S^{n-1}$.*

Clearly every individual term separately satisfies $\|\boldsymbol{x}\|_K^2 \leq r^2$ and $\|\boldsymbol{x}\|_{K^\circ}^2 \leq r^2$ for $\boldsymbol{x} \in S^{n-1}$ — so the non-trivial claim is that their sum can never exceed r^2+1.

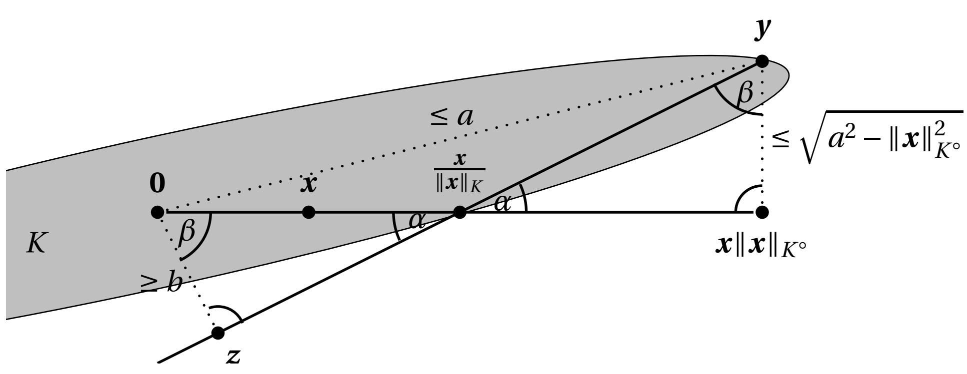
We will now prove a more general version. It will be worth keeping the following pictures in mind:



$\frac{1}{a}\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq \frac{1}{b}\|\boldsymbol{x}\|_2$ $\qquad$ $b\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_{K^\circ} \leq a\|\boldsymbol{x}\|_2$

Lemma 7.9 (Distance Lemma)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex symmetric set with $\frac{1}{a}\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq \frac{1}{b}\|\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Suppose that there is an $\boldsymbol{x} \in S^{n-1}$ with $\left(b\|\boldsymbol{x}\|_K\right)^2 + \left(\frac{\|\boldsymbol{x}\|_{K^\circ}}{a}\right)^2 = t > 1$, then $\frac{a}{b} \leq \frac{1}{t-1}$.*

Proof. Fix the point $\boldsymbol{x} \in S^{n-1}$ from the assumption and let $\boldsymbol{y}$ be the dual point, that means $\|\boldsymbol{x}\|_{K^\circ} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ and $\|\boldsymbol{y}\|_K = 1$. Note that by assumption $\|\boldsymbol{y}\|_2 \leq a\|\boldsymbol{y}\|_K = a$. We will argue only using points in the 2-dimensional plane $\mathrm{span}\{\boldsymbol{x}, \boldsymbol{y}\}$. Consider the line through the points $\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_K}$ and $\boldsymbol{y}$ and denote the point on that line that minimizes the $\|\cdot\|_2$ norm by $\boldsymbol{z} = (1+\lambda)\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2} - \lambda\boldsymbol{y}$ for some $\lambda \in \mathbb{R}$. Note that $\|\boldsymbol{z}\|_2^2 = (1+\lambda)^2 - \frac{2\lambda(1+\lambda)}{\|\boldsymbol{x}\|_2}\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \lambda^2\|\boldsymbol{y}\|_2^2$ and from $\langle \boldsymbol{x}, \boldsymbol{y}\rangle > 0$ we can see that $\lambda > 0$. We reproduce the figure of the book [AAGM15] in a modified form:



Next, note that $\|\boldsymbol{z}\|_K = \|(1+\lambda)\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_K} - \lambda\boldsymbol{y}\|_K \geq (1+\lambda) - \lambda\|\boldsymbol{y}\|_K = (1+\lambda) - \lambda = 1$ and by assumption $\|\boldsymbol{z}\|_2 \geq b\|\boldsymbol{z}\|_K \geq b$. Note that we have two triangles in the picture that are similar as their angles are both $\alpha, \beta, 90^o$. Then the ratio of the length of the two sides incident to the β-angle are identical, i.e.

$$\begin{aligned}
b\|\boldsymbol{x}\|_K \quad & \overset{\|\boldsymbol{z}\|_2 \geq b}{\leq} \quad \frac{\|\boldsymbol{z}\|_2}{1/\|\boldsymbol{x}\|_K} = \frac{\|\boldsymbol{z}-\boldsymbol{0}\|_2}{\|\boldsymbol{0}-\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_K}\|_2} \overset{\text{triangle similarity}}{=} \frac{\|\boldsymbol{y}-\|\boldsymbol{x}\|_{K^\circ}\boldsymbol{x}\|_2}{\|\boldsymbol{y}-\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_K}\|_2} \\
& \overset{\text{Phytagoras}}{=} \frac{\sqrt{\|\boldsymbol{y}\|_2^2-\|\boldsymbol{x}\|\boldsymbol{x}\|_{K^\circ}\|_2^2}}{\sqrt{(\|\boldsymbol{y}\|_2^2-\|\boldsymbol{x}\|\boldsymbol{x}\|_{K^\circ}\|_2^2)+\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_K}-\boldsymbol{x}\|\boldsymbol{x}\|_{K^\circ}\|_2^2}} \\
& \overset{\|\boldsymbol{y}\|_2 \leq a}{\leq} \frac{\sqrt{a^2-\|\boldsymbol{x}\|_{K^\circ}^2}}{\sqrt{a^2-\|\boldsymbol{x}\|_{K^\circ}^2+(\frac{1}{\|\boldsymbol{x}\|_K}-\|\boldsymbol{x}\|_{K^\circ}^2)^2}} \\
& = \frac{\sqrt{a^2-\|\boldsymbol{x}\|_{K^\circ}}}{\sqrt{a^2-2\frac{\|\boldsymbol{x}\|_{K^\circ}}{\|\boldsymbol{x}\|_K}+\frac{1}{\|\boldsymbol{x}\|_K^2}}}
\end{aligned}$$

Squaring and multiplying both sides by $\frac{1}{a^2\|\boldsymbol{x}\|_K^2}$ gives

$$\frac{b^2}{a^2} \leq \frac{1-(\|\boldsymbol{x}\|_{K^\circ}/a)^2}{a^2\|\boldsymbol{x}\|_K^2-2\|\boldsymbol{x}\|_K\|\boldsymbol{x}\|_{K^\circ}+1}$$

Multiplying by the right hand side denominator and adding $\frac{2\|\boldsymbol{x}\|_K\|\boldsymbol{x}\|_{K^\circ}-a^2\|\boldsymbol{x}\|_K^2}{(a/b)^2}$ to both sides gives

$$\begin{aligned}
\frac{b^2}{a^2} \quad & \leq \quad 1-\underbrace{\left(\left(\frac{\|\boldsymbol{x}\|_{K^\circ}}{a}\right)^2+\left(b\|\boldsymbol{x}\|_K\right)^2\right)}_{=t}+\frac{2b}{a}(b\|\boldsymbol{x}\|_K)\frac{\|\boldsymbol{x}\|_K^\circ}{a} \\
& \overset{AMGM}{\leq} \quad 1-t+\frac{b}{a}\underbrace{\left(\left(b\|\boldsymbol{x}\|_K\right)^2+\left(\frac{\|\boldsymbol{x}\|_{K^\circ}}{a}\right)^2\right)}_{=t} = 1-t+\frac{tb}{a}
\end{aligned}$$

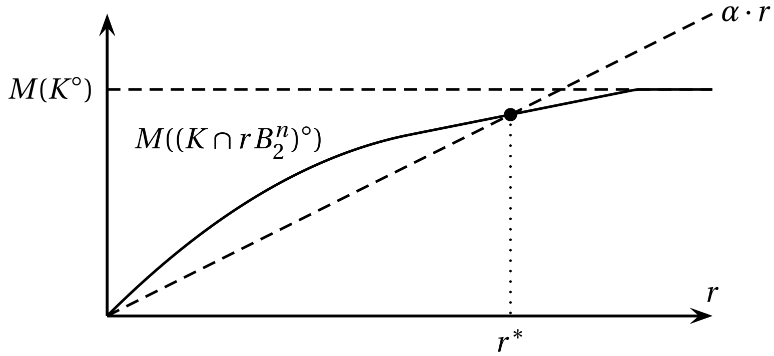
where we use the AMGM inequality of $\alpha\cdot\beta \leq \frac{1}{2}(\alpha^2+\beta^2)$. Rearranging gives

$$0 \leq 1-t+\frac{tb}{a}-\frac{b^2}{a^2} = \underbrace{\left(1-\frac{b}{a}\right)}_{>0}\cdot\left(\frac{b}{a}-(t-1)\right)$$

hence one must have $\frac{b}{a}-(t-1) \geq 0$ which is $\frac{a}{b} \leq \frac{1}{t-1}$. □

7.5.1 Radius of random sections — with a twist

Take a symmetric convex body $K \subseteq \mathbb{R}^n$ and consider the quantity $M((K \cap rB_2^n)^\circ)$. We know that this is a proxy for the radius of a ball that has the same average width as $K \cap rB_2^n$. For $r \to \infty$ we expect that $M((K \cap rB_2^n)^\circ) \to M(K^\circ)$ while for very small r, one has $M(K \cap rB_2^n)^\circ) = M((rB_2^n)^\circ) = r$. Hence for any $0 < \alpha < 1$, there will be a value $r^* > 0$ so that $M((K \cap r^* B_2^n)^\circ) = \alpha \cdot r^*$.



We can prove an interesting lemma where we get the intersection with rB_2^n for free — however, we seem to have little control on how large or small that radius r is going to be. In the statement of the lemma we use the unspecified constant $C_{\mathrm{PT}} \geq 1$ from the Pajor-Tomczak Theorem (Theorem 7.2) which guarantees that $\Pr[\mathrm{radius}(K \cap F) \leq \frac{C_{\mathrm{PT}}}{\sqrt{1-\lambda}} M(K^\circ)] \geq 1 - e^{-\Theta(1-\lambda)\cdot n}$ for a random λn-dimensional subspace. It was in fact proven by Gordon [Gor88] that the constant C_{PT} can be taken arbitrarily close to 1 if on the other hand n is at least a threshold that depends on the target choice of C_{PT} and λ. We will later take the liberty to pick C_{PT} to be rather close to 1, without proving Gordon's Theorem.

Lemma 7.10. *Fix $0 < \lambda < 1$ and let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Let $r > 0$ be a value so that $M((K \cap rB_2^n)^\circ) = \frac{9}{10} \frac{\sqrt{1-\lambda}}{C_{\mathrm{PT}}} \cdot r$. Then*

$$\Pr[\mathit{radius}(K \cap F) < r] \geq 1 - 4\exp(-\Theta(1-\lambda) \cdot n)$$

where $F \subseteq \mathbb{R}^n$ is a random λn-dimensional subspace.

Proof. We apply the Pajor-Tomczak Theorem to $K \cap rB_2^n$ and with probability $1 - 4\exp(-\Theta(1-\lambda) \cdot n)$, a random λn-dimensional subspace F will satisfy

$$\mathrm{radius}((K \cap rB_2^n) \cap F) \overset{\text{Thm. 7.2}}{\leq} \frac{C_{\mathrm{PT}}}{\sqrt{1-\lambda}} \cdot M((K \cap rB_2^n)^\circ) \overset{\text{choice of } r}{=} \frac{9}{10} r$$

Expressed in terms of norms, the subspace F satisfies

$$\max\left\{\|\boldsymbol{x}\|_K, \frac{1}{r}\|\boldsymbol{x}\|_2\right\} = \|\boldsymbol{x}\|_{K \cap rB_2^n} \geq \frac{10}{9} \cdot \frac{1}{r} \cdot \|\boldsymbol{x}\|_2 \quad \forall \boldsymbol{x} \in F$$

But that means there is no point $\boldsymbol{x} \in F$ where this maximum is attained for the second term — hence $\|\boldsymbol{x}\|_K \geq \frac{10}{9} \cdot \frac{1}{r} \|\boldsymbol{x}\|_2$ for every $\boldsymbol{x} \in F$. And that shows the claim. $\square$

Note that the choice of the constant $\frac{9}{10}$ was somewhat arbitrary.

7.5.2 The Linear Duality Theorem

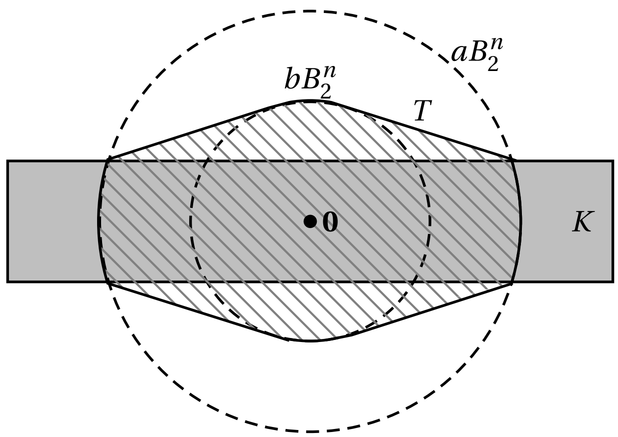
Now we have everything for the final result of this chapter:

Theorem 7.11 (Milman [Mil90b]). *Let $K \subseteq \mathbb{R}^n$ be a convex symmetric set. Then there exists a subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) \geq \Omega(n)$ so that either radius$(K \cap F) \leq O(1)$ or radius$(K^\circ \cap F) \leq O(1)$.*

Proof. Let $0 < \lambda < 1$ be a small constant that we determine later. Let $a > 0$ be a value so that $M((K \cap aB_2^n)^\circ) = \frac{\sqrt{1-\lambda}}{2C_{\text{PT}}} \cdot a$. Also let $b > 0$ be the value so that $M((K^\circ \cap \frac{1}{b}B_2^n)^\circ) = \frac{\sqrt{1-\lambda}}{2C_{\text{PT}}} \cdot \frac{1}{b}$. The crucial part of the proof is proving that a and b are not far apart:

Claim. *One has $\frac{b}{a} \leq 10$.*

Proof of claim. Consider the set $T := \text{conv}((K \cap aB_2^n) \cup bB_2^n))$. Note that by construction $bB_2^n \subseteq T \subseteq aB_2^n$.



We know that

$$M(T^\circ) \overset{T \supseteq K \cap aB_2^n}{\geq} M((K \cap aB_2^n)^\circ) \overset{\text{choice of } a}{=} \frac{\sqrt{1-\lambda}}{2C_{\text{PT}}} a \quad (*)$$

Moreover we can use the fact that $(A \cap B)^\circ = \text{conv}(A^\circ \cup B^\circ)$ in $(***)$ to obtain

$$M(T) \overset{T \subseteq \text{conv}(K \cup bB_2^n)}{\geq} M(\text{conv}(K \cup bB_2^n)) \overset{(***)}{=} M\Big(\Big(K^\circ \cap \frac{1}{b}B_2^n\Big)^\circ\Big) \overset{\text{choice of } b}{=} \frac{\sqrt{1-\lambda}}{2C_{\text{PT}}} \cdot \frac{1}{b}. \quad (**)$$

Then we can estimate that

$$(bM(T))^2 + \Big(\frac{1}{a}M(T^\circ)\Big)^2 \overset{(*)+(**)}{\geq} \Big(b\frac{9}{10}\frac{\sqrt{1-\lambda}}{C_{\text{PT}}} \cdot \frac{1}{b}\Big)^2 + \Big(\frac{1}{a} \cdot \frac{9}{10}\frac{\sqrt{1-\lambda}}{C_{\text{PT}}} a\Big)^2 = \frac{81}{50}\frac{1-\lambda}{C_{PT}^2} \geq \frac{6}{5}$$

if we choose $C_{PT} = \frac{11}{10}$ and $\lambda := \frac{1}{10}$.

That means the premise of the Distance Lemma is satisfied for the two *averages* $M(T)$ and $M(T^\circ)$. But it is not hard to extract a concrete point $\boldsymbol{x}^*$ that satisfies the premise as well. As $bB_2^n \subseteq T$, we know that the map $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|_T$ is b-Lipschitz and hence for a suitable choice of constant, the set $A := \{\boldsymbol{x} \in S^{n-1} \mid \|\boldsymbol{x}\|_T \geq M(T) - \Theta(\frac{b}{\sqrt{n}})\}$ has measure $\sigma(A) \geq \frac{3}{4}$. Similarly, $\frac{1}{a} \subseteq T^\circ$ and so $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|_{T^\circ}$ is $\frac{1}{a}$-Lipschitz which implies that $B := \{\boldsymbol{x} \in S^{n-1} \mid \|\boldsymbol{x}\|_{T^\circ} \geq M(T^\circ) - \Theta(\frac{1}{a\sqrt{n}})$ also has $\sigma(B) \geq \frac{3}{4}$. Then any $\boldsymbol{x}^* \in A \cap B$ satisfied $(b\|\boldsymbol{x}^*\|_T)^2 + (\frac{1}{a}\|\boldsymbol{x}^*\|_{T^\circ})^2 \geq \frac{6}{5} - \Theta(\frac{1}{\sqrt{n}}) \geq 1 + \frac{1}{10}$ for n large enough. Then by the Distance Lemma, we conclude that $\frac{a}{b} \leq 10$. $\square$

Now we go back to reasoning about the original body K. Consider a random λn-dimensional subspace $F \subseteq \mathbb{R}^n$. We will do a case split, dependent on whether the radius a is large or small.

- *Case* $a \leq 1$. Then we use the choice of a and apply Lemma 7.10 with body K and $r := a$ to get $\text{radius}(K \cap F) \leq a \leq 1$.

- *Case* $a > 1$. Crucially in this case one has $b \geq \frac{1}{10}$. We apply Lemma 7.10 to K° with $r = \frac{1}{b}$ and obtain that with probability at least $1 - 4\exp(-\Theta((1-\lambda)n))$ one has that $\text{radius}(K^\circ \cap F) < \frac{1}{b} \leq 10$.

$\square$

7.6 Exercises

Exercise 7.1.

(i) What upper bound on $\text{radius}(B_\infty^n \cap A)$ does the Pajor-Tomczak Theorem provide where A is a uniform k-dimensional subspace.

(ii) What upper bound on $\text{radius}(B_1^n \cap A)$ does the Pajor-Tomczak Theorem provide where A is a uniform k-dimensional subspace.

(iii) Consider the body $K := rB_\infty^n$ for some $r > 0$. For which range of r is $\text{radius}(K \cap A) \leq O(1)$ (for most subspaces A with $\dim(A) = \frac{n}{2}$; using the estimate of (i))? For which range of r is $\text{radius}(K^\circ \cap A) \leq O(1)$ (again for most subspaces with $\dim(A) = \frac{n}{2}$ and using the estimate from (ii))?

Exercise 7.2.

(i) Prove that for any subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) = k$ one has $\text{radius}(B_\infty^n \cap F) \geq \sqrt{k}$.

(ii) Prove that there are constants $c, c' > 0$ so that for all n there exists a subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) \geq cn$ so that $\text{radius}(B_1^n \cap F) \leq \frac{c'}{\sqrt{n}}$.
Hint. Use the linear duality theorem (Theorem 7.11).

Chapter 8

M-ellipsoids and applications

One of the most powerful results in convex geometry is the existence of the *M-ellipsoid*. It implies that for many purposes of volume computation and covering numbers, any symmetric convex body can be replaced with an ellipsoid while only incurring a constant factor loss.

Formally, for a symmetric convex body $K \subseteq \mathbb{R}^n$ and a universal constant $C > 0$, we say that an origin-centered ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ is an *M-ellipsoid for K*, if $\text{Vol}_n(K) = \text{Vol}_n(\mathcal{E})$ and

$$C^{-n} \leq \frac{\text{Vol}_n(K+P)}{\text{Vol}_n(\mathcal{E}+P)} \leq C^n \quad \text{and} \quad C^{-n} \leq \frac{\text{Vol}_n(K^\circ+P)}{\text{Vol}_n(\mathcal{E}^\circ+P)} \leq C^n$$

for every symmetric convex body $P \subseteq \mathbb{R}^n$. In this chapter we will show the result of Milman that indeed every symmetric convex body K admits such an ellipsoid. We will prove that such an ellipsoid $\mathcal{E}$ has the property that the covering numbers $N(K,\mathcal{E}), N(\mathcal{E},K), N(K^\circ,\mathcal{E}^\circ)$ and $N(\mathcal{E}^\circ,K^\circ)$ are all upper bounded by $2^{\Theta(n)}$. In fact, the reverse holds in the sense that if all these covering numbers are upper-bounded by $2^{O(n)}$ and $\text{Vol}_n(\mathcal{E}) = \text{Vol}_n(K)$, then $\mathcal{E}$ is an M-ellipsoid for K.

For the existence of the M-ellipsoid, we will use the method of *isotrophic symmetrization*. The same method will also enable us to give a more intuitive proof of the *Reverse Santaló-Inequality* due to Bourgain-Milman that we have seen in Chapter 7. Recall that combined we know that:

> *Blaschke-Santaló-Bourgain-Milman Theorem. For every symmetric convex body K one has*
>
> $$C^n \leq \frac{Vol_n(K) \cdot Vol_n(K^\circ)}{Vol_n(B_2^n)^2} \leq 1$$
>
> *where $C > 0$ is a universal constant. The lower bound also holds also for any asymmetric body with $\mathbf{0} \in int(K)$.*

Finally we will be able to proof that for any centrally symmetrix convex body K there is a linear transformation $U : \mathbb{R}^n \to \mathbb{R}^n$ with $|\det(\boldsymbol{U})| = 1$ so that the Brunn-Minkowski inequality with $U(K)$ is appproximately tight.

8.1 M-position and equivalences

For the time being, we will restrict our attention to symmetric convex sets and later in Chapter 8.7 discuss which results extend to the asymmetric case. It was useful for example in Chapter 3 to introduce the notion of *John's position* for a convex body. Similarly we want to introduce the following terminology:

Definition 8.1. We say that a centrally symmetric convex body $K \subseteq \mathbb{R}^n$ with $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(rB_2^n)$ is in *M-position with constant C*, if rB_2^n is the M-ellipsoid for K, i.e. if for every symmetric convex body $P \subseteq \mathbb{R}^n$ one has

$$C^{-n} \leq \frac{\mathrm{Vol}_n(K+P)}{\mathrm{Vol}_n(rB_2^n + P)} \leq C^n \quad \text{and} \quad C^{-n} \leq \frac{\mathrm{Vol}_n(K^\circ + P)}{\mathrm{Vol}_n\left(\frac{1}{r}B_2^n + P\right)} \leq C^n$$

In particular we will later show that for every symmetric convex body K, there is a linear transformation A so that $A(K)$ is in M-position.

Remark 1. We should point out that alternatively in the definition of an M-ellipsoid $\mathcal{E}$ we could have required that P may be an arbitrary convex body (not necessarily symmetric). But these two candidate definitions are equivalent. To see this recall that $N(P-P,P) \leq 2^{5n}$ by Lemma 4.8. Then $2^{-5n}\mathrm{Vol}_n(K+(P-P)) \leq \mathrm{Vol}_n(K+P) \leq \mathrm{Vol}_n(K+(P-P))$ for any two convex bodies K, P where K is symmetric.

A crucial fact is that K and $\mathcal{E}$ have similar volume expansion, if and only if one can cover one with few copies of the other. Note that in the following Theorem one may have a loss of constants when moving between the statements $(A), (B), (C)$. Recall that $N(K, \mathcal{E})$ is the minimum number of translates of $\mathcal{E}$ to cover K.

Theorem 8.2 (Equivalence for M-ellipsoids)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $\mathcal{E} \subseteq \mathbb{R}^n$ be an origin centered ellipsoid with $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(\mathcal{E})$. Then the following is equivalent*

(A) One has

$$C^{-n} \leq \frac{\mathrm{Vol}_n(K+P)}{\mathrm{Vol}_n(\mathcal{E}+P)} \leq C^n \quad \text{and} \quad C^{-n} \leq \frac{\mathrm{Vol}_n(K^\circ+P)}{\mathrm{Vol}_n(\mathcal{E}^\circ+P)} \leq C^n$$

for a universal constant C and every symmetric convex body $P \subseteq \mathbb{R}^n$.

(B) *One has*

$$\max\{N(K,\mathcal{E}), N(\mathcal{E},K), N(K^\circ,\mathcal{E}^\circ), N(\mathcal{E}^\circ,K^\circ)\} \le 2^{\Theta(n)}.$$

(C) *One has* $N(K,\mathcal{E}) \le 2^{O(n)}$.

Proof. After applying a linear transformation, we may assume that $\mathcal{E} = B_2^n = \mathcal{E}^\circ$. Then (A) is equivalent to the property that K is in M-position (with respect to B_2^n).

$(A) \Rightarrow (B)$. From Chapter 4 we know that the covering number is bounded by volume ratios and

$$N(K,B_2^n) \overset{\text{Lem 4.3}}{\le} \frac{\mathrm{Vol}_n(K+\frac{1}{2}B_2^n)}{\mathrm{Vol}_n(\frac{1}{2}B_2^n)} \overset{(A)}{\le} 2^{\Theta(n)} \frac{\mathrm{Vol}_n(B_2^n+\frac{1}{2}B_2^n)}{\mathrm{Vol}_n(\frac{1}{2}B_2^n)} \le 2^{\Theta(n)}.$$

The other 3 covering estimates work along the same lines.

$(B) \Rightarrow (A)$. Suppose that $N := N(K,B_2^n) \le 2^{\Theta(n)}$. Then there is a covering $K \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + B_2^n)$ which means that for a convex body T one has $K + T \subseteq \bigcup_{i=1}^N (\boldsymbol{x}_i + B_2^n) + T$. Hence $\mathrm{Vol}_n(K+T) \le 2^{O(n)} \cdot \mathrm{Vol}_n(B_2^n + T)$. Again, the other 3 cases work along the same lines.

$(B) \Rightarrow (C)$. Obvious.

$(C) \Rightarrow (B)$. This requires some non-trivial work. In particular we require the *Duality of Covering Numbers* that we have proven in Theorem 7.7: *For every pair $K,T \subseteq \mathbb{R}^n$ of symmetric, convex bodies one has* $2^{-\Theta(n)} N(T^\circ,K^\circ) \le N(K,T) \le 2^{\Theta(n)} N(T^\circ,K^\circ)$. From the assumption $N(K,B_2^n) \le 2^{O(n)}$, we then get $N(B_2^n,K^\circ) \le 2^{O(n)}$. We can the 3rd covering number by applying the volume ratio argument

$$\begin{aligned}
N(K^\circ,B_2^n) &\overset{\text{Lem 4.3}}{\le} \frac{\mathrm{Vol}_n(K^\circ+\frac{1}{2}B_2^n)}{\mathrm{Vol}_n(\frac{1}{2}B_2^n)} \le 2^{O(n)} \cdot \frac{\mathrm{Vol}_n(K^\circ+B_2^n)}{\mathrm{Vol}_n(B_2^n)} \\
&\overset{\text{Blaschke-Santaló}}{\le} 2^{O(n)} \cdot \frac{\mathrm{Vol}_n(B_2^n)}{\mathrm{Vol}_n((K^\circ+B_2^n)^\circ)} \\
&\overset{K^\circ+B_2^n \subseteq 2\mathrm{conv}(K^\circ \cup B_2^n)}{\le} 2^{O(n)} \frac{\mathrm{Vol}_n(B_2^n)}{\mathrm{Vol}_n((\mathrm{conv}(K^\circ \cup B_2^n))^\circ)} \\
&\overset{\mathrm{conv}(A\cup B)^\circ = A^\circ \cap B^\circ}{=} 2^{O(n)} \cdot \frac{\mathrm{Vol}_n(B_2^n)}{\mathrm{Vol}_n(K \cap B_2^n)} \overset{(*)}{\le} 2^{O(n)} \cdot \underbrace{N(K,B_2^n)}_{\le 2^{O(n)}} \cdot \underbrace{\frac{\mathrm{Vol}_n(B_2^n)}{\mathrm{Vol}_n(K)}}_{=1}
\end{aligned}$$

In $(*)$ we have used that $\mathrm{Vol}_n(K \cap B_2^n) \cdot N(K,B_2^n) \ge \mathrm{Vol}_n(K)$ as the intersection volume $\mathrm{Vol}_n((\boldsymbol{x}+B_2^n) \cap K)$ is maximized for $\boldsymbol{x} = \mathbf{0}$. The last of the 4 estimates on $N(B_2^n,K)$ follows again from the duality of covering numbers. ∎

We note that the equivalence with (C) requires the Duality of Entropy numbers (Theorem 7.7) which uses the Bourgain-Milman inequality (Theorem 7.6), which in turn we had proven using the Quotient of Subspaces Theorem. In this chapter we will see a more intuitive proof for the Bourgain-Milman inequality, hence we will refrain from using (C) before finishing the reasoning of that alternative proof.

Again, by Remark 1 the equivalences in Theorem 8.2 are also true if P is any (potentially asymmetric) convex body.

8.2 Isomorphic symmetrization

Now we come to a procedure that is called *isomorphic symmetrization* which we have seen in a variant in Theorem 7.11. This procedure will be the key in proving the existence of an M-ellipsoid. The idea is to take an arbitrary symmetric convex body K, intersect it with a large ball and add in the convex hull with a smaller ball. The consequence is that we will obtain another symmetric convex body K_1 that is "more round". To get some intuition about the choice of parameters suppose that K is in ℓ*-position* and we have scaled K so that $M(K) = 1$ and $M(K^\circ) \leq O(\log n)$. Then if we apply the procedure with parameter $\alpha := 1$, then it means we intersect K with a ball of radius $M(K^\circ) \leq O(\log n)$ and we take the convex hull with the ball of radius 1. In turn we will end up with a body K that has volume $2^{-\Theta(n)} \leq \mathrm{Vol}_n(K_1)/\mathrm{Vol}_n(K) \leq 2^{\Theta(n)}$ while its geometric distance to the Euclidean ball has dropped to $d_G(K_1, B_2^n) \leq O(\log n)$. Note that the choice of n^{100} in the following assumption $\frac{1}{n^{100}} B_2^n \subseteq K \subseteq n^{100} B_2^n$ is somewhat arbitrary — any polynomial relation to B_2^n would suffice.

Lemma 8.3. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body $\frac{1}{n^{100}} B_2^n \subseteq K \subseteq n^{100} B_2^n$ so that $M(K) = 1$ and let $1 \leq \alpha \leq n^{1/3}$. Define an inner radius of $R_{in} := \frac{1}{\alpha}$ and an outer radius of $R_{out} := M(K^\circ) \cdot \alpha$ and consider*

$$K_1 := conv\Big((K \cap R_{out} B_2^n) \cup R_{in} B_2^n\Big)$$

Then the following holds:

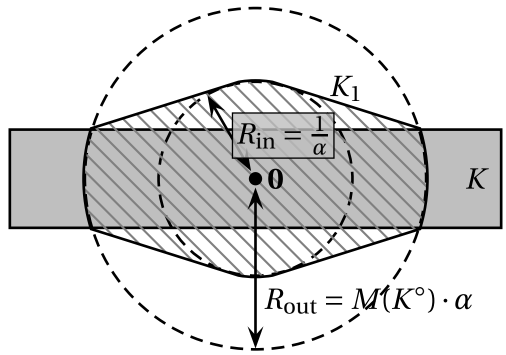
(1) One has

$$e^{-\Theta(n/\alpha^2)} \leq \frac{Vol_n(K_1)}{Vol_n(K)} \leq e^{\Theta(n/\alpha^2)} \quad \text{and} \quad e^{-\Theta(n/\alpha^2)} \leq \frac{Vol_n(K_1^\circ)}{Vol_n(K^\circ)} \leq e^{\Theta(n/\alpha^2)}$$

(2) For every symmetric convex body $P \subseteq \mathbb{R}^n$ one has

$$\exp(-\Theta(n/\alpha^2)) \leq \frac{Vol_n(K_1 + P)}{Vol_n(K + P)} \leq \exp(\Theta(n/\alpha^2))$$

(3) *It holds that* $d_G(K_1, B_2^n) \leq M(K) \cdot M(K^\circ) \cdot \alpha^2$.



Proof. Note that since $M(K) = 1$ one has $M(K^\circ) \geq 1$, and hence indeed $R_{\text{in}} \leq R_{\text{out}}$. Next, it is clear that the "moreover" part is true as $d_G(K_1, B_2^n) \leq R_{\text{out}}/R_{\text{in}} = M(K) \cdot M(K^\circ) \cdot \alpha^2$. The idea behind the main part of the analysis is that K_1 is sandwiched between Euclidean balls of radius proportional to $M(K)$ and $M(K^\circ)$ and hence with Sudakov's primal and dual Inequality we can obtain estimates on how many copies a ball takes to cover K; similar for the polar. Then from those covering numbers we get an estimate on the volume ratios.

Claim I. *One has* $N(K, R_{out}B_2^n), N(R_{in}B_2^n, K), N(K^\circ, \frac{1}{R_{in}}B_2^n), N(\frac{1}{R_{out}}B_2^n, K^\circ) \leq \exp(\Theta(n/\alpha^2))$.

Proof of claim II. We begin by proving the 2 estimates for the covering numbers involving K. First, by Sudakov's Inequality (Theorem 4.12; recall that $w(K) = 2 \cdot M(K^\circ)$), we only need few copies of the outer ball to cover K:

$$N(K, R_{\text{out}}B_2^n) \leq \exp\Big(\Theta(n) \cdot \Big(\underbrace{\frac{M(K^\circ)}{R_{\text{out}}}}_{=1/\alpha}\Big)^2\Big) = \exp(\Theta(n/\alpha^2)) \quad (*)$$

Secondly, by the dual Sudakov Inequality (Theorem 4.10) one has

$$N(R_{\text{in}}B_2^n, K) = N(B_2^n, \alpha K) \leq \exp\Big(\Theta(n) \cdot \Big(\frac{M(K)}{\alpha}\Big)^2\Big) = e^{\Theta(n/\alpha^2)}$$

Now we discuss the polar. Again with Sudakov's Inequality

$$N\Big(K^\circ, \frac{1}{R_{\text{in}}}B_2^n\Big) \leq \exp\Big(\Theta(n) \cdot \Big(\frac{M(K)}{1/R_{\text{in}}}\Big)^2\Big) = e^{\Theta(n/\alpha^2)}$$

and again with the dual Sudakov Inequality

$$N\Big(\frac{1}{R_{\text{out}}}B_2^n, K^\circ\Big) = N(B_2^n, R_{\text{out}}K^\circ) \leq \exp\Big(\Theta(n) \cdot \Big(\frac{M(K^\circ)}{R_{\text{out}}}\Big)^2\Big) = e^{\Theta(n/\alpha^2)} \quad \square$$

Claim II. *One has* $e^{-\Theta(n/\alpha^2)} \leq \frac{Vol_n(K_1)}{Vol_n(K)} \leq e^{\Theta(n/\alpha^2)}$.

Proof of Claim II. We use the bounds on the covering numbers to obtain estimates on $\mathrm{Vol}_n(K_1)$. First, we obtain a lower bound on the volume of K_1 by

$$\mathrm{Vol}_n(K_1) \geq \mathrm{Vol}_n(K \cap R_{\mathrm{out}} B_2^n) \geq \frac{\mathrm{Vol}_n(K)}{N(K, R_{\mathrm{out}} B_2^n)} \stackrel{\text{Claim I}}{\geq} \exp(-\Theta(n/\alpha^2)) \cdot \mathrm{Vol}_n(K)$$

For the upper bound we estimate that

$$\begin{aligned} \mathrm{Vol}_n(K_1) &\leq \mathrm{Vol}_n\Big(\mathrm{conv}\Big(K \cup R_{\mathrm{in}} B_2^n\Big)\Big) \\ &\leq O(n) \cdot n^{100} \cdot N(R_{\mathrm{in}} B_2^n, K) \cdot \mathrm{Vol}_n(K) \leq e^{\Theta(n/\alpha^2)} \cdot \mathrm{Vol}_n(K) \end{aligned}$$

where we absorb the polynomial factor into the term $e^{\Theta(n/\alpha^2)} \geq e^{\Theta(n^{1/3})}$ as $\alpha \leq n^{1/3}$. Here we use that for symmetric bodies K, L with $L \subseteq \beta K$ one has $\mathrm{Vol}_n(\mathrm{conv}(K \cup L)) \leq O(\beta n) \cdot N(L, K) \cdot \mathrm{Vol}_n(K)$. Note that in our case $R_{\mathrm{in}} B_2^n \subseteq n^{100} K$ as $\alpha \geq 1$. □

Claim III. *One has* $e^{-\Theta(n/\alpha^2)} \leq \frac{Vol_n(K_1^\circ)}{Vol_n(K^\circ)} \leq e^{\Theta(n/\alpha^2)}$.

Proof of Claim III. Now we discuss the volume for the polar. First, the polar has the form

$$K_1^\circ = (K \cap R_{\mathrm{out}} B_2^n)^\circ \cap \Big(R_{\mathrm{in}} B_2^n\Big)^\circ = \mathrm{conv}\Big(K^\circ \cup \frac{1}{R_{\mathrm{out}}} B_2^n\Big) \cap \frac{1}{R_{\mathrm{in}}} B_2^n$$

Then

$$\mathrm{Vol}_n(K_1^\circ) \geq \mathrm{Vol}_n\Big(K^\circ \cap \frac{1}{R_{\mathrm{in}}} B_2^n\Big) \geq \frac{\mathrm{Vol}_n(K^\circ)}{N(K^\circ, \frac{1}{R_{\mathrm{in}}} B_2^n)} \stackrel{\text{Claim I}}{\geq} e^{-\Theta(n/\alpha^2)} \mathrm{Vol}_n(K^\circ)$$

Similar to before, we obtain the upper bound

$$\mathrm{Vol}_n(K_1^\circ) \leq \mathrm{Vol}_n\Big(\mathrm{conv}\Big(K^\circ \cup \frac{1}{R_{\mathrm{out}}} B_2^n\Big)\Big) \leq O(n) \cdot \frac{n^{100}}{R_{\mathrm{out}}} \cdot \underbrace{N\Big(\frac{1}{R_{\mathrm{out}}} B_2^n, K^\circ\Big)}_{\leq e^{\Theta(n/\alpha^2)}} \cdot \mathrm{Vol}_n(K^\circ)$$

Again we use $\frac{1}{n^{100}} B_2^n \subseteq K^\circ$. □

Claim IV. *One has* $e^{-\Theta(n/\alpha^2)} \leq \frac{Vol_n(K_1+P)}{Vol_n(K+P)} \leq e^{\Theta(n/\alpha^2)}$ *for any symmetric convex body* P.

Proof of Claim IV. We will crucially use two lemmas on covering numbers and volumes from Chapter 4. First we write

$$\begin{aligned} \mathrm{Vol}_n(K_1 + P) &\leq \mathrm{Vol}_n\Big(\mathrm{conv}\Big(K \cup R_{\mathrm{in}} B_2^n\Big) + P\Big) \\ &\stackrel{\text{Lem 4.16}}{\leq} O(n^{101}) \cdot N\Big(R_{\mathrm{in}} B_2^n, K\Big) \cdot \mathrm{Vol}_n(K + P) \\ &\leq e^{O(n/\alpha^2)} \cdot \mathrm{Vol}_n(K + P) \end{aligned}$$

Here we use that $K \subseteq n^{100} B_2^n$ and we absorb the polynomial factor into the exponential one. Conversely,

$$\begin{aligned} \mathrm{Vol}_n(K_1 + P) &= \mathrm{Vol}_n\Big(\mathrm{conv}\Big((K \cap R_{\mathrm{out}} B_2^n) \cup R_{\mathrm{in}} B_2^n\Big) + P\Big) \\ &\geq \mathrm{Vol}_n((K \cap R_{\mathrm{out}} B_2^n) + P) \\ &\overset{\text{Lem 4.15}}{\geq} \frac{\mathrm{Vol}_n(K+P)}{N(K, R_{\mathrm{out}} B_2^n)} \\ &\geq e^{-\Theta(n/\alpha^2)} \cdot \mathrm{Vol}_n(K+P) \quad \square \end{aligned}$$

This concludes the Theorem. □

For the sake of completeness, we state the consequence for a body that is not already in M-position:

Corollary 8.4. Let $K \subseteq \mathbb{R}^n$ be any symmetric convex body let $1 \leq \alpha \leq n^{1/3}$ be a parameter. Then there exists a symmetric convex body $K_1 \subseteq \mathbb{R}^n$ so that the following holds:

(1) One has

$$\frac{\mathrm{Vol}_n(K_1)}{\mathrm{Vol}_n(K)}, \frac{\mathrm{Vol}_n(K_1^\circ)}{\mathrm{Vol}_n(K^\circ)}, \frac{\mathrm{Vol}_n(K_1+P)}{\mathrm{Vol}_n(K+P)} \in \Big[e^{-\Theta(n/\alpha^2)}, e^{\Theta(n/\alpha^2)}\Big]$$

for every symmetric convex body $P \subseteq \mathbb{R}^n$.

(2) One has $d_{BM}(K_1) \leq O(\alpha^2 \cdot \log(d_{BM}(K)+1))$.

Proof. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map so that $A(K)$ is in ℓ-position, scaled so that $M(A(K)) = 1$. Recall that due to Theorem 6.19 this means that $M(A(K)) \cdot M(A(K)^\circ) \leq C_0 \ln(d_{BM}(K)+1)$ for some constant C_0. Moreover, if we have a body in ℓ-position scaled so that $M(A(K)) = 1$, then from Cor 5.10 we know that $\Theta(\frac{1}{\sqrt{n}}) B_2^n \subseteq A(K)$ and $\Theta(\frac{1}{\sqrt{n}\log(n)}) B_2^n \subseteq A(K)^\circ$. Then Theorem 8.3 can be applied and (by a slight abuse of notation) we obtain a body that we write as $A(K_1)$ so that $e^{-\Theta(n/\alpha^2)} \leq \frac{\mathrm{Vol}_n(A(K_1))}{\mathrm{Vol}_n(A(K))} \leq e^{\Theta(n/\alpha^2)}$ holds and $d_{BM}(K_1) = d_{BM}(A(K_1)) \leq C_0 \ln(d_{BM}(K)+1)$ is true. Importantly, a linear map does not change the ratio of volumes, hence also $\frac{\mathrm{Vol}_n(K_1)}{\mathrm{Vol}_n(K)} = \frac{\mathrm{Vol}_n(A(K_1))}{\mathrm{Vol}_n(A(K))}$. The same reasoning holds for the other two properties. □

8.3 Iterated Isomorphic Symmetrization

The next obvious step is to iteratively apply the Isomorphic Symmetrization to a body K until we reach a body $\tilde{K}$ with $d_{BM}(\tilde{K}) \leq O(1)$. We can summarize the properties as follows:

Theorem 8.5. *Let $K \subseteq \mathbb{R}^n$ be any symmetric convex body. Then there exists a symmetric convex body $\tilde{K} \subseteq \mathbb{R}^n$ with $d_{BM}(\tilde{K}) \leq O(1)$ so that*

$$\frac{Vol_n(\tilde{K})}{Vol_n(K)}, \frac{Vol_n(\tilde{K}^\circ)}{Vol_n(K^\circ)}, \frac{Vol_n(\tilde{K}+P)}{Vol_n(K+P)} \in \left[2^{-\Theta(n)}, 2^{\Theta(n)}\right]$$

for every symmetric convex body $P \subseteq \mathbb{R}^n$.

Proof. Let $C_0 > 0$ be the constant from Cor 8.4 so that $d_{BM}(K_1) \leq C_0\alpha^2 \cdot \log(d_{BM}(K)+1)$. Consider the following iterative procedure:

(1) Set $K_0 := K$
(2) FOR $t = 0$ TO ∞ DO
    (3) IF $d_{BM}(K_t) \leq C_1 := 100C_0^3$ then set $T := t$ and return K_T
    (4) Apply Isomorphic Symmetrization of Cor 8.4 to K_t with parameter $\alpha_t := (d_{BM}(K_t))^{1/4}$ and let K_{t+1} be the outcome.

We know that the Isomorphic Symmetrization decreases the Banach Mazur distance to the Euclidean ball to

$$d_{BM}(K_{t+1}) \overset{\substack{\text{Cor 8.4+}\\ \text{choice of } \alpha_t}}{\leq} C_0 \ln(d_{BM}(K_t)+1) \cdot \sqrt{d_{BM}(K_t)} \leq \frac{1}{16} d_{BM}(K_t)$$

As this distance is strictly decreasing, the procedure will terminate and the final iterate satisfies $d_{BM}(K_T) \leq C_1$. Moreover by the choice of the parameter α_t, we conclude that $\alpha_{t+1} \leq \frac{1}{2}\alpha_t$ where the final parameter satisfies $\alpha_{T-1} = (d_{BM}(K_{T-1}))^{1/4} \geq 1$. The bounds on the loss in volume from Lemma 8.3 imply that in each iteration $t \in \{0, \dots, T-1\}$ we have

$$e^{-\Theta(n/\alpha_t^2)} \leq \frac{\mathrm{Vol}_n(K_{t+1})}{\mathrm{Vol}_n(K_t)} \leq e^{\Theta(n/\alpha_t^2)}$$

Then accounting the loss of volume over all iterations we get

$$\frac{\mathrm{Vol}_n(\tilde{K})}{\mathrm{Vol}_n(K)} = \prod_{t=0}^{T-1} \frac{\mathrm{Vol}_n(K_{t+1})}{\mathrm{Vol}_n(K_t)} \geq \prod_{t=0}^{T-1} e^{-\Theta(n/\alpha_t^2)} = \exp\Big(-\Theta(n) \cdot \underbrace{\frac{1}{\alpha_{T-1}^2}}_{\leq 1} \cdot \underbrace{\sum_{k=0}^{T-1} (2^{-k})^2}_{\leq 2}\Big) \geq C_1^{-n} e^{-\Theta(n)}$$

The upper bound can be obtained similarly; the same holds for the other two ratios $\mathrm{Vol}_n(\tilde{K}^\circ)/\mathrm{Vol}_n(K^\circ)$ and $\mathrm{Vol}_n(\tilde{K}+P)/\mathrm{Vol}_n(K+P)$. $\square$

Note that actually $O(\log^*(n))$ many iterations suffice if one makes the choice of $\alpha_t := \ln(d_{BM}(K_t)+1)$. However, the analysis is somewhat simpler with our choice. The ellipsoid that is a $O(1)$-approximation to $\tilde{K}$ will indeed be the M-ellipsoid of K. In order to prove this, we first show the Bourgain-Milman Inequality.

8.4 Another proof for the Bourgain-Milman Inequality

Recall the notation $s(K) := \mathrm{Vol}_n(K)\cdot\mathrm{Vol}_n(K^\circ)$ and that the Blaschke-Santaló inequality says that $s(K)\le s(B_2^n)$ for any symmetric convex body and we have seen the reverse approximate inequality by Bourgain and Milman in Chapter 7. Here we can present an alternative proof that gives both directions of the inequality at once and also provides a more intuitive explanation why the inequality holds.

Theorem 8.6 (Bourgain-Milman Inequality)**.** *Consider any symmetric convex body $K\subseteq\mathbb{R}^n$. Then one has $(\frac{1}{C})^n \le \frac{s(K)}{s(B_2^n)} \le C^n$, for some constant $C>0$.*

Proof. We would like to remind the reader that for any bijective linear map $A:\mathbb{R}^n\to\mathbb{R}^n$ one has $s(A(K)) = s(K)$ as volume ratios do not change under a linear map since $\mathrm{Vol}_n(A(K)) = |\det(\boldsymbol{A})|\cdot\mathrm{Vol}_n(K)$ and $\mathrm{Vol}_n(A(K)^\circ) = \frac{1}{|\det(\boldsymbol{A})|}\cdot\mathrm{Vol}_n(K^\circ)$. Now, apply the Iterated Isomorphic Symmetrization of Theorem 8.5 to the body K and let $\tilde{K}$ be the result. As $d_{BM}(\tilde{K})\le O(1)$, we know that $d_G(\tilde{K}, A(B_2^n))\le O(1)$ for some linear map. Then

$$\begin{aligned} s(K) &= \mathrm{Vol}_n(K)\cdot\mathrm{Vol}_n(K^\circ) \stackrel{\text{Theorem 8.5}}{\ge} 2^{-\Theta(n)}\cdot\mathrm{Vol}_n(\tilde{K})\cdot\mathrm{Vol}_n(\tilde{K}^\circ) \\ &\ge 2^{-\Theta(n)}\cdot 2^{-\Theta(n)}\cdot\mathrm{Vol}_n(A(B_2^n))\cdot\mathrm{Vol}_n(A(B_2^n)^\circ) = 2^{-\Theta(n)}\cdot 2^{-\Theta(n)}\cdot s(B_2^n) \end{aligned}$$

the upper bound follows along the same lines. □

8.5 Existence of M-ellipsoids

Now back to proving the existence of M-ellipsoids.

Theorem 8.7 (Milman [Mil88])**.** *Every symmetric convex body $K\subseteq\mathbb{R}^n$ has an M-ellipsoid $\mathcal{E}_K$.*

Proof. Again, we apply the Iterated Isomorphic Symmetrization from Theorem 8.5 to K and denote $\tilde{K}$ as outcome. Let $\mathcal{E}$ be an ellipsoid with $\mathrm{Vol}_n(\mathcal{E}) = \mathrm{Vol}_n(K)$ and

$\frac{1}{C}\mathcal{E} \subseteq \tilde{K} \subseteq C \cdot \mathcal{E}$ for some constant C. Then

$$N(K,\mathcal{E}) \overset{\text{Lem 4.3}}{\leq} \frac{\mathrm{Vol}_n(K+\frac{1}{2}\mathcal{E})}{\mathrm{Vol}_n(\frac{1}{2}\mathcal{E})} \leq 2^{\Theta(n)} \cdot \frac{\mathrm{Vol}_n(\tilde{K}+\frac{1}{2}\mathcal{E})}{\mathrm{Vol}(\frac{1}{2}\mathcal{E})} \leq 2^{\Theta(n)} \frac{\mathrm{Vol}_n(C\cdot\mathcal{E}+\frac{1}{2}\mathcal{E})}{\mathrm{Vol}_n(\frac{1}{2}\mathcal{E})} = 2^{\Theta(n)} \cdot (2C+1)^n$$

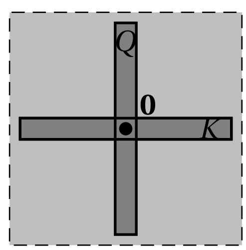
As $C = O(1)$, this means $N(K,\mathcal{E}) \leq 2^{\Theta(n)}$. Now we can apply Theorem 8.2.(C) and infer that $\mathcal{E}$ is indeed an M-ellipsoid for K. This reasoning has indeed used the Bourgain-Milman inequality. ☐

8.6 The Reverse Brunn-Minkowski Inequality

Recall that the Brunn-Minkowski inequality says that for any measurable sets $K, Q \subseteq \mathbb{R}^n$ one has

$$\mathrm{Vol}_n(K+Q)^{1/n} \geq \mathrm{Vol}_n(K)^{1/n} + \mathrm{Vol}_n(Q)^{1/n}$$

It is not hard to come up with examples of even symmetric convex sets where this inequality is arbitrarily weak and $\mathrm{Vol}_n(K+Q)$ is a lot higher than the guarantee:



But it turns out that if both sets are in M-position, then the inequality is indeed close to tight.

Theorem 8.8 (Reverse Brunn-Minkowski Inequality — Milman [Mil86])**.** *For every symmetric convex body K, there is an invertible linear map $U_K : \mathbb{R}^n \to \mathbb{R}^n$ with $|\det(\boldsymbol{U}_K)| = 1$ so that the following holds: for every pair $K, Q \subseteq \mathbb{R}^n$ of symmetric convex bodies one has*

$$Vol_n\Big(U_K(K) + t \cdot U_Q(Q)\Big)^{1/n} \leq O(1) \cdot \Big(Vol_n(U_K(K))^{1/n} + t \cdot Vol_n(U_Q(Q))^{1/n}\Big) \quad \forall t \geq 0$$

Proof. We use the proof following [Mil88]. Let $\mathcal{E}_K$ be the M-ellipsoid for body K. We pick the map U_K so that $U_K(\mathcal{E}_K) = r_K \cdot B_2^n$ where r_K is the radius with $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(r_K B_2^n)$. In other words, $U_K(K)$ is in M-position. Note that the maps U_K do not change volume and in particular $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(U_K(K))$. Let C be the constant from the M-ellipsoid definition. Then

$$\begin{aligned}
\mathrm{Vol}_n\Big(U_K(K) + t\cdot U_Q(Q)\Big)^{1/n} &\overset{M\text{-ellipsoid for } K}{\leq} C\cdot \mathrm{Vol}_n\Big(r_K B_2^n + t\cdot U_Q(Q)\Big)^{1/n} \\
&\overset{M\text{-ellipsoid for } Q}{\leq} C^2\cdot \mathrm{Vol}_n\Big(r_K B_2^n + t\cdot r_Q B_2^n\Big)^{1/n} \\
&= C^2\cdot \Big(\mathrm{Vol}_n(r_K B_2^n)^{1/n} + t\cdot \mathrm{Vol}_n(r_Q B_2^n)^{1/n}\Big) \\
&= C^2\cdot \Big(\mathrm{Vol}_n(K)^{1/n} + t\cdot \mathrm{Vol}_n(Q)^{1/n}\Big)
\end{aligned}$$

□

8.7 Extension to the non-symmetric case

In this section, we discuss how some of the result that we have proven for symmetric convex sets, also extend to non-symmetric sets. First of all, the inequality of Bourgain-Milman also applies to the non-symmetric case:

Theorem 8.9 (Bourgain-Milman Inequality for Asymmetric Convex Sets)**.** *For any convex body K containing $\mathbf{0}$ in the interior one has*

$$\frac{Vol_n(K)\cdot Vol_n(K^\circ)}{Vol_n(B_2^n)^2} \geq 2^{-\Theta(n)}$$

Proof. We define $T := K - K$ which is a symmetric convex body with $T \supseteq K$. Then using the inequality of *Rogers-Shephard* we know that $\mathrm{Vol}_n(K-K) \leq 4^n\cdot \mathrm{Vol}_n(K)$. As $T^\circ \subseteq K^\circ$ we obtain

$$\underbrace{\mathrm{Vol}_n(K)}_{\geq 4^{-n}\mathrm{Vol}_n(T)}\cdot \underbrace{\mathrm{Vol}_n(K^\circ)}_{\geq \mathrm{Vol}_n(T^\circ)} \geq 4^{-n}\cdot \mathrm{Vol}_n(T)\cdot \mathrm{Vol}_n(T^\circ) \overset{\substack{\text{Bourgain-Milman}\\ \text{with } T \text{ symmetric}}}{\geq} 2^{-\Theta(n)}\cdot \mathrm{Vol}_n(B_2^n)^2$$

□

One should mention that the reverse of this inequality (that means the Blaschke-Santaló inequality) does not necessarily hold. Consider $K := [-1, M] \subseteq \mathbb{R}$ with M large. Then $K^\circ = [-1, \frac{1}{M}]$ and $\mathrm{Vol}_1(K)\cdot \mathrm{Vol}_1(K^\circ) = (M-1)\cdot(1-\frac{1}{M}) \geq M/2$ for $M \geq 4$. Next, we can prove that the powerful duality result for covering numbers from Theorem 7.7 extends to asymmetric sets:

Theorem 8.10. *Let $K, T \subseteq \mathbb{R}^n$ be convex bodies, both having $\mathbf{0}$ as barycenter. Then*

$$2^{-\Theta(n)} N(T^\circ, K^\circ) \le N(K,T) \le 2^{\Theta(n)} N(T^\circ, K^\circ)$$

Proof. It suffices to prove one direction. We will use that we know the inequality for symmetric sets already from the König-Milman Theorem (Theorem 7.7). Then

$$\begin{aligned}
N(T^\circ, K^\circ) &\overset{\text{monotonicity}}{\le} N(\mathrm{conv}(T^\circ \cup (-T)^\circ), K^\circ \cap (-K)^\circ) \\
&\overset{\text{König-Milman}}{\le} 2^{O(n)} \cdot N\big((K^\circ \cap (-K)^\circ)^\circ, (\mathrm{conv}(T^\circ \cup (-T)^\circ))^\circ\big) \\
&= 2^{O(n)} \cdot N\big(\mathrm{conv}(K \cup (-K)), T \cap (-T)\big) \\
&\le 2^{O(n)} \cdot \underbrace{N(K-K,K)}_{\le 2^{O(n)}} \cdot N(K,T) \cdot \underbrace{N(T, T\cap(-T))}_{\le 2^{O(n)}} \\
&\overset{\text{RS+MP}}{\le} 2^{O(n)} \cdot N(K,T)
\end{aligned}$$

where in the last step we use Rogers-Shepphard (Theorem 1.47) and Milman-Pajor (Theorem 4.5). □

In order to obtain an M-ellipsoid for a non-symmetric body K, the obvious strategy is to try either the M-ellipsoid for larger symmetric set $K-K$ or the one for the smaller symmetric set $K \cap (-K)$. Luckily in terms of M-position it does not matter which one we use:

Lemma 8.11. *Let $K \subseteq \mathbb{R}^n$ be a convex body with barycenter at $\mathbf{0}$. Then $K-K$ is in M-position if and only if $K \cap (-K)$ is in M-position (possibly with different constants).*

Proof. Follows from $K \cap (-K) \subseteq K-K$ and $N(K-K, K\cap(-K)) \le 2^{5n}$, see Lemma 4.8. Then the covering numbers $N(K-K, rB_2^n)$ and $N(K\cap(-K), rB_2^n)$ differ by at most a factor of $2^{O(n)}$. □

Finally we can argue that after a translation, every convex set admits an M-ellipsoid.

Theorem 8.12 (M-ellipsoids for asymmetric convex sets)**.** *For any convex body $K \subseteq \mathbb{R}^n$ with barycenter at $\mathbf{0}$, there exists an ellipsoid $\mathcal{E}$ so that*

(A) One has

$$C^{-n} \le \frac{\mathrm{Vol}_n(K+P)}{\mathrm{Vol}_n(\mathcal{E}+P)} \le C^n \quad \text{and} \quad C^{-n} \le \frac{\mathrm{Vol}_n(K^\circ+P)}{\mathrm{Vol}_n(\mathcal{E}^\circ+P)} \le C^n$$

for every symmetric convex body P.

(B) One has

$$\max\{N(K,\mathcal{E}), N(K^\circ,\mathcal{E}^\circ), N(\mathcal{E},K), N(\mathcal{E}^\circ,K^\circ)\} \leq 2^{O(n)}$$

Proof. Applying a linear transformation to K does not change the validity of the statement and it also leaves the barycenter at the origin. Hence we may assume that the symmetric body $K-K$ is in M-position with respect to the ball B_2^n. We claim that then the properties hold for the choice of $\mathcal{E} := B_2^n$. As $K-K$ is in M-position we know that there is a constant C_0 so that

$$\frac{\mathrm{Vol}_n(K+P)}{\mathrm{Vol}_n(B_2^n+P)} \leq \frac{\mathrm{Vol}_n((K-K)+P)}{\mathrm{Vol}_n(B_2^n+P)} \leq C_0^n$$

for every symmetric convex body P. For the lower bound, suppose that $K\cap(-K)$ is in M-position with respect to the ball $r_{\mathrm{int}}B_2^n$ and with respect to the constant C_1. Then in particular we have $\mathrm{Vol}_n(r_{\mathrm{int}}B_2^n) = \mathrm{Vol}_n(K\cap(-K)) \geq 4^{-n}\mathrm{Vol}_n(K-K) = 8^{-n}\mathrm{Vol}_n(B_2^n)$ using the Milman-Pajor Theorem (Theorem 4.5 with Cor 4.6) from which we conclude that $\frac{1}{4} \leq r_{\mathrm{int}} \leq 1$. Then

$$\frac{\mathrm{Vol}_n(K+P)}{\mathrm{Vol}_n(B_2^n+P)} \geq \frac{\mathrm{Vol}_n((K\cap(-K))+P)}{\mathrm{Vol}_n(B_2^n+P)} \geq (1/4)^n\frac{\mathrm{Vol}_n((K\cap(-K))+P)}{\mathrm{Vol}_n(r_{\mathrm{int}}B_2^n+P)} \geq (1/4)^n C_1^n$$

The volume bounds for the polar work similar, using the inclusions $(K-K)^\circ \subseteq K^\circ$ and $(K\cap(-K))^\circ \supseteq K^\circ$. For the covering estimates we have $N(K,B_2^n) \leq N(K-K,B_2^n) \leq 2^{O(n)}$ and $N(B_2^n,K) \leq N(B_2^n,K\cap(-K)) \leq 2^{O(n)}$. The other two estimates follow from duality of covering numbers. $\square$

8.8 Regular M-ellipsoids

We want to briefly mention a powerful extension to the concept of M-ellipsoids. If $\mathcal{E}$ is the M-ellipsoid for K, then we know that in particular $N(K,\mathcal{E}) \leq 2^{O(n)}$. But this alone does not necessarily guarantee that the covering number is shrinking substantially if the ellipsoid is scaled, i.e. it is not clear how large $N(K,t\mathcal{E})$ with $t>1$ would need to be. A stronger form of so-called *regular M-ellipsoids* provides such guarantees! exactly. Details can be found in the textbooks of Artstein-Avidan et al [AAGM15] or Pisier [Pis89b] or in the original work of Pisier [Pis89a].

Theorem 8.13 (Pisier)**.** *Let $0 \leq p < 2$ and let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then there exists an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ so that*

$$N(K,t\mathcal{E}), N(\mathcal{E},tK), N(K^\circ,t\mathcal{E}^\circ), N(\mathcal{E}^\circ,tK^\circ) \leq \exp\left(Cn/t^p\right) \quad \forall t \geq 1,$$

where $C := C(p) > 0$ is a constant only depending on p.

Note that setting $t := (2Cn)^{1/p}$ we have $N(K, t\mathcal{E}), N(\mathcal{E}, tK) = 1$ and hence $\frac{1}{t}\mathcal{E} \subseteq K \subseteq t\mathcal{E}$. From this one can see that the statement must be false for $p > 2$. With some extra work, the concept of regular M-ellipsoids can be generalized to non-symmetric convex bodies (where the constant of $\frac{2}{5}$ seems unlikely to be tight).

Theorem 8.14 (Vritsiou [Vri23])**.** *Let $0 \leq p < \frac{2}{5}$ and let $K \subseteq \mathbb{R}^n$ be a convex body with $\mathbf{0}$ as barycenter. Then there exists an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ so that*

$$N(K, t\mathcal{E}), N(\mathcal{E}, tK), N(K^\circ, t\mathcal{E}^\circ), N(\mathcal{E}^\circ, tK^\circ) \leq \exp\left(Cn/t^p\right) \quad \forall t \geq 1,$$

where $C := C(p) > 0$ is a constant only depending on p.

We note that an earlier paper by Klartag and Milman [KM05] (Prop 2.2) already claimed the same result, but the authors later report that the proof is flawed. However, parts of the proof strategy of Klartag and Milman turned out to be sound. We refer to the work of Vritsiou [Vri23] for details.

8.9 Exercises

Exercise 8.1.
For any symmetric convex set $K \subseteq \mathbb{R}^n$ there is a linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ so that one has $\max\left\{N(A(K), Q), N(Q, A(K)), N(A(K)^\circ, Q), N(Q, A(K)^\circ)\right\} \leq 2^{O(n)}$ where $Q := \frac{1}{\sqrt{n}} B_\infty^n$ is the cube of side length $\frac{2}{\sqrt{n}}$.

Exercise 8.2.
Let $K \subseteq B_2^n$ be a symmetric convex set contained in the unit ball. Then there exists an M-ellipsoid $\mathcal{E}^*$ of K with $\mathcal{E}^* \subseteq CB_2^n$ where C is a universal constant.

Exercise 8.3.
There is a universal constant $C > 0$ so that the following holds for even n. Let $K \subseteq B_2^n$ be a symmetric convex body and let $0 < \delta \leq 1$ be so that $\mathrm{Vol}_n(K) = \delta^n \mathrm{Vol}_n(B_2^n)$. Then there is a subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) = n/2$ so that $\mathrm{Vol}_{n/2}((\frac{C}{\delta^2} K \cap F) \cap B_2^n) \geq 2^{-Cn} \mathrm{Vol}_n(B_2^n)$.
Hint. Consider the M-ellipsoid for K and show that at least $n/2$ of the axes have length at least $\Theta(\delta^2)$. Choose F as the span of the $n/2$ longest axis. You may use the result from the previous exercise.

Chapter 9

The Gaussian Approach

The goal of this chapter will be to develop a more fine-grained understanding of the *Gaussian mean width*

$$g(T) := \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}\left[\sup_{\boldsymbol{t}\in T}\langle \boldsymbol{t},\boldsymbol{x}\rangle\right]$$

for a set $T \subseteq \mathbb{R}^n$. Recall that this is not really a new quantity as $g(T) = \frac{a_n}{2}\cdot w(T)$ where $w(T)$ is the mean width of T and $a_n = \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_2]$ is the expected length of a Gaussian where $\sqrt{n}\cdot\sqrt{\frac{n}{n+1}} \leq a_n \leq \sqrt{n}$, see Lemma 1.1.

So far we have mostly considered sets T that were convex bodies, but convexity is not needed for the definition to make sense. In fact, we can even restrict to the finite case whenever it is helpful:

Lemma 9.1. *Let $T \subseteq \mathbb{R}^n$. If T is unbounded, then $g(T) = \infty$. If T is bounded, then for any $\varepsilon > 0$ there is a finite set $T' \subseteq T$ with $g(T) - \varepsilon \leq g(T') \leq g(T)$.*

We leave the proof as an exercise. Throughout this chapter we use $\log(x) := \ln(x)$. Whenever we state an upper bound including a $O(\log(N))$-term the reader should interpret this as $O(\log(2N))$ to cover the case $N = 1$.

9.1 Gaussian Random Processes

It will turn out to be useful to discuss a more general concept. Unlike the other chapters in this set of lecture notes, for the remainder of this chapter we will follow the exposition in the recent textbook by Vershynin [Ver19]. First we want to introduce Gaussians that are not necessarily "standard". If $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ is a symmetric, positive-definite matrix, then we write $N(\boldsymbol{0}, \boldsymbol{M})$ as the distribution of a Gaussian random vector $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{M})$ with expectation $\mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{M})}[\boldsymbol{x}] = \boldsymbol{0}$ and covariance matrix $\mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{M})}[\boldsymbol{x}\boldsymbol{x}^T] = \boldsymbol{M}$. Equivalently, for $i, j \in [n]$ one has $\mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{M})}[x_i \cdot x_j] = M_{ij}$. We remind the reader that a random Gaussian $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$ can be generated by sampling the coordinates $x_1, \dots, x_n \sim N(0,1)$ independently. We start with the following rather abstract definition:

Definition 9.2. A *random process* is a collection of random variables $(X_t)_{t\in T}$ on the same probability space. A random process $(X_t)_{t\in T}$ is a *Gaussian process*, if for every $a : T \to \mathbb{R}$ with finite support, there is a $\sigma \geq 0$ with $\sum_{t\in T} a_t X_t \sim N(0, \sigma^2)$. For such a Gaussian process we define a *metric* d by $d(s,t) := \mathbb{E}[|X_s - X_t|^2]^{1/2}$ for $s, t \in T$.

For this definition we make no restrictions on the index set T. We would like to mention that all Gaussians for the rest of this chapter are "centered", which means they have mean 0.

We give a name to the type of Gaussian process that corresponds to our geometric setting:

Definition 9.3. For a set $T \subseteq \mathbb{R}^n$, the random process $(X_{\boldsymbol{t}})_{\boldsymbol{t}\in T}$ with $X_t = \langle \boldsymbol{x}, \boldsymbol{t} \rangle$ for $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$ is called a *canonical Gaussian process.*

If T is finite then we call $(X_t)_{t\in T}$ a *finite* Gaussian process. We summarize a few facts about Gaussians:

Lemma 9.4. *The following holds:*

(A) *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two finite Gaussian processes with identical covariance matrices (i.e. $\mathbb{E}[X_t X_s] = \mathbb{E}[Y_t Y_s]$ for $t, s \in T$). Then the distributions of $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ are identical.*

(B) *Let $(X_t)_{t\in T}$ be a Gaussian process and let $T' \subseteq T$ be a finite set with $\mathbb{E}[X_t X_s] = 0$ for all $t, s \in T'$ with $t \neq s$. Then the random variables $(X_t)_{t\in T'}$ are independent.*

(C) *For $s > 0$ one has $\frac{1}{\sqrt{2\pi}}(\frac{1}{s} - \frac{1}{s^3})e^{-s^2/2} \leq \Pr_{x\sim N(0,\sigma^2)}[x \geq s\cdot\sigma] \leq \frac{1}{\sqrt{2\pi}s} e^{-s^2/2}$*

Typically (A) is proven by arguing that the Fourier transforms of both processes are identical and then conclude that the distributions themselfs are identical. We skip the proof though.

It is useful to observe that the distance metric for a canonical Gaussian process $(X_{\boldsymbol{t}})_{\boldsymbol{t}\in T}$ corresponds to the Euclidean distance of the vectors:

$$d(\boldsymbol{s}, \boldsymbol{t}) = \mathbb{E}\left[\left|X_{\boldsymbol{s}} - X_{\boldsymbol{t}}\right|^2\right]^{1/2} = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}\left[\langle \boldsymbol{x}, \boldsymbol{s} - \boldsymbol{t}\rangle^2\right]^{1/2} = \|\boldsymbol{s} - \boldsymbol{t}\|_2 \quad \forall \boldsymbol{s}, \boldsymbol{t} \in T$$

We will often work with a general Gaussian process as it gives us some more flexibility — in some sense it is "dimension-free". But it is useful to keep in mind that at least for the finite case both concepts are equivalent:

Lemma 9.5. *A Gaussian process $(X_t)_{t\in T}$ with finite T is always a canonical Gaussian process.*

Proof sketch. Let $\boldsymbol{M} \in \mathbb{R}^{T\times T}$ be the covariance matrix of the Gaussian process, meaning that $M_{s,t} = \mathbb{E}[X_s X_t]$ for $s,t \in T$. Then $\boldsymbol{M}$ is positive semidefinite and so there are vectors $\{\boldsymbol{u}_t\}_{t\in T}$ with $\boldsymbol{u}_t \in \mathbb{R}^T$ so that $M_{s,t} = \langle \boldsymbol{u}_s, \boldsymbol{u}_t\rangle$. Now consider the random process $(Y_t)_{t\in T}$ with $Y_t := \langle \boldsymbol{x}, \boldsymbol{u}_t\rangle$ with $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I}_{|T|})$. Then $\mathbb{E}[Y_s Y_t] = \langle \boldsymbol{u}_s, \boldsymbol{u}_t\rangle = M_{st}$ for all $s,t \in T$, meaning that the Gaussian processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ have the same covariance matrix. Then by Lemma 9.4.(A) the random processes have the same distribution. ☐

Now we come back to our main task of understanding the *expected supremum of a Gaussian process* which is

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big].$$

First, we want to emphasize that for independent Gaussians, the expected supremum is rather easy to understand:

Lemma 9.6 (Geometric version of Lemma 9.7)**.** *Let $T = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ be pairwise orthogonal vectors with $A \leq \|\boldsymbol{x}_i\|_2 \leq B$ for all $i \in [N]$. Then*

$$\Theta\big(A\sqrt{\log(N)}\big) \leq g(T) \leq \Theta\big(B\sqrt{\log(N)}\big).$$

Moreover, the upper bound also holds for non-orthogonal vectors.

We will prove this in the setting of Gaussian processes:

Lemma 9.7. *Let $(X_t)_{t\in T}$ be a finite Gaussian process where (i) one has $A \leq \mathbb{E}[X_t^2]^{1/2} \leq B$ for $t \in T$ and (ii) one has $\mathbb{E}[X_{t_1} X_{t_2}] = 0$ for all $t_1 \neq t_2$. Then*

$$\Theta\Big(A\sqrt{\log(|T|)}\Big) \leq \mathbb{E}\Big[\sup_{t\in T} X_t\Big] \leq \Theta\Big(B\sqrt{\log(|T|)}\Big)$$

and the upper bound holds without (ii).

Proof. It suffices to show the claim for $|T|$ large enough. We also may assume that for some $t_0 \in T$ the lower bound on the variance is tight, i.e. $A = \mathbb{E}[X_{t_0}^2]^{1/2}$.

Note that the assumptions tell us that (i) the standard deviation of the random Gaussians X_t is between A and B and (ii) the Gaussians X_t are independent.

For the upper bound we simply observe that $\|X_t\|_{\psi_2} \leq O(B)$ and so the claim follows from Lemma 3.17.(i).

Now we come to the lower bound. Again using the estimate from Lemma 9.4.(C) one can verify that for some small constant $c_2 > 0$ and $\lambda := c_2 A\sqrt{\log(|T|)}$ one has $\Pr[X_t \geq \lambda] \geq \frac{1}{|T|}$ for each $t \in T$. Using the independence gives

$$\begin{aligned}\Pr[\exists t \in T : X_t \geq \lambda] &\stackrel{\text{indep}}{=} 1 - \prod_{t\in T}\Big(1 - \underbrace{\Pr[X_t \geq \lambda]}_{\geq 1/|T|}\Big) \\ &\geq 1 - \Big(1 - \frac{1}{|T|}\Big)^{|T|} \stackrel{1-x\geq e^{-x/2}\forall 0\leq x\leq \frac{1}{2}}{\geq} 1 - e^{-|T|\cdot\frac{|T|}{2}} \geq \frac{1}{4}\end{aligned}$$

We are almost done with the analysis, but there is the slight issue that the random variable $\sup_{t\in T} X_t$ might also be negative sometimes. But the negative contribution can be loosely bounded by $|\mathbb{E}[\min\{0, \sup_{t\in T} X_t\}]| \leq |\mathbb{E}[\min\{0, X_{t_0}\}]| \leq \mathbb{E}[|X_{t_0}|] = \sqrt{\frac{2}{\pi}} \cdot A$. Then

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big] \geq \lambda \cdot \Pr[\exists t \in T : X_t \geq \lambda] - \sqrt{\frac{2}{\pi}}A \geq c_3 A\sqrt{\log(|T|)}$$

as claimed. □

This leaves us with two main questions:

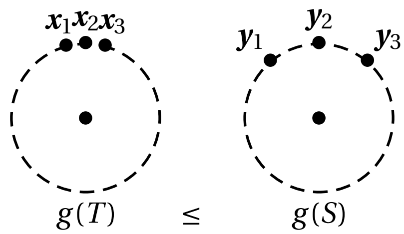
1. How can we show a *lower bound* on $\mathbb{E}[\sup_{t\in T} X_t]$ without orthogonality/independence?

2. How can we get better a better *upper bound* than just using a union bound.

9.2 Slepian's Inequality

We will now develop a tool to answer the first question. In a simple to state form, we can prove the following:

Lemma 9.8 (Sudakov-Fernique Inequality — Geometric version)**.** *Let $T = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ and $S = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_N\}$ be two sets of vectors with $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2$ for all $i, j \in [N]$. Then $g(T) \leq g(S)$.*



The idea behind the proof is to continuously "morph" the set T into the set S and prove that along the way, the expected supremum is non-decreasing. Again we use the probabilistic view for the proof.

9.2.1 Gaussian Interpolation

We show a few lemmas first that are fairly specific about Gaussians.

Lemma 9.9. *Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Then*

$$\mathop{\mathbb{E}}_{X \sim N(0,1)}[f'(X)] = \mathop{\mathbb{E}}_{X \sim N(0,1)}[X \cdot f(X)]$$

Proof. Let $\gamma(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ be the density function of the Gaussian and w.l.o.g. assume that $f(x) \neq 0 \Rightarrow a < x < b$ for some bounded interval $I = [a, b] \subseteq \mathbb{R}$. Then for $X \sim N(0,1)$ one has[1]

$$\begin{aligned} \mathbb{E}[f'(X)] &= \int_I f'(x) \cdot \gamma(x) dx \overset{\text{partial int.}}{=} \underbrace{\left[f(x) \cdot \gamma(x)\right]_a^b}_{=0} - \int_I f(x) \cdot \gamma'(x) dx \\ &\overset{(*)}{=} -\int_I f(x) \cdot (-x \cdot \gamma(x))\, dx = \mathbb{E}[X \cdot f(X)] \end{aligned}$$

where we use in $(*)$ that $\gamma'(x) = -x \cdot \gamma(x)$ as one can easily verify. □

Here is the multi-dimensional analogue (we skip the proof):

Lemma 9.10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and let $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ be a PSD matrix. Then for $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{M})$ one has*

$$\underbrace{\mathbb{E}[\boldsymbol{X} \cdot f(\boldsymbol{X})]}_{\in \mathbb{R}^n} = \underbrace{\boldsymbol{M}}_{\in \mathbb{R}^{n \times n}} \underbrace{(\mathbb{E}[\nabla f(\boldsymbol{X})])}_{\in \mathbb{R}^n}$$

[1] Recall that *integration by parts* gives $\int_a^b f'(x)\gamma(x)dx + \int_a^b f(x)\gamma'(x)dx = f(x) \cdot \gamma(x)|_a^b = 0$ as $f(a) = 0 = f(b)$.

Note that writing this equation coordinate-wise gives

$$\mathop{\mathbb{E}}_{\boldsymbol{X}}[X_i \cdot f(\boldsymbol{X})] = \sum_{j=1}^{n} M_{ij} \mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[\frac{\partial f}{\partial x_j}(\boldsymbol{X})\Big]$$

again with $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{M})$.

Now we come to the *Gaussian interpolation principle*. The idea is that we consider two Gaussian random vectors $\boldsymbol{X}, \boldsymbol{Y}$ and a smooth interpolation $\boldsymbol{Z}(u)$ between those where $\boldsymbol{Z}(0) = \boldsymbol{Y}$ and $\boldsymbol{Z}(1) = \boldsymbol{X}$. Then we can characterize exactly how $\mathbb{E}[f(\boldsymbol{Z}(u))]$ changes as u increases. For example suppose that $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{M}^{\boldsymbol{X}})$ and $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{M}^{\boldsymbol{Y}})$ are Gaussians where the covariance matrices $\boldsymbol{M}^{\boldsymbol{X}}$ and $\boldsymbol{M}^{\boldsymbol{Y}}$ are identical, except for a single entry (i, j) where $M_{ij}^X = M_{ij}^Y + \varepsilon$. Then indeed

$$\frac{d}{du}\mathbb{E}[f(\boldsymbol{Z}(u))] = \frac{\varepsilon}{2} \cdot \mathbb{E}\Big[\frac{\partial f^2}{\partial x_i \partial x_j}\big(\boldsymbol{Z}(u)\big)\Big]$$

Lemma 9.11 (Gaussian Interpolation)**.** *Let $\boldsymbol{M}^{\boldsymbol{X}}, \boldsymbol{M}^{\boldsymbol{Y}} \in \mathbb{R}^{n\times n}$ be PSD matrices and let $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{M}^{\boldsymbol{X}})$ and $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{M}^{\boldsymbol{Y}})$ be independent Gaussian random vectors. Define the interpolation*

$$\boldsymbol{Z}(u) := \sqrt{u}\cdot \boldsymbol{X} + \sqrt{1-u}\cdot \boldsymbol{Y} \quad \text{for } u \in [0,1].$$

Then for any twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ one has

$$\frac{d}{du}\mathbb{E}[f(\boldsymbol{Z}(u))] = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(M_{ij}^X - M_{ij}^Y)\cdot \mathbb{E}\Big[\frac{\partial f^2}{\partial x_i \partial x_j}\big(\boldsymbol{Z}(u)\big)\Big]$$

Proof. Using the chain rule[2] we can write

$$\begin{aligned}
\frac{d}{du}\mathbb{E}[f(\boldsymbol{Z}(u))] &\overset{\text{chain rule}}{=} \sum_{i=1}^{n}\mathbb{E}\Big[\frac{\partial f}{\partial x_i}\big(\boldsymbol{Z}(u)\big)\cdot \frac{dZ_i}{du}\Big] \\
&\overset{(*)}{=} \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\Big[\frac{\partial f}{\partial x_i}(\boldsymbol{Z}(u))\cdot\Big(\frac{X_i}{\sqrt{u}} - \frac{Y_i}{\sqrt{1-u}}\Big)\Big] \\
&\overset{\text{Claim I+II}}{=} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(M_{ij}^X - M_{ij}^Y)\cdot \mathbb{E}\Big[\frac{\partial f^2}{\partial x_i \partial x_j}\big(\boldsymbol{Z}(u)\big)\Big]
\end{aligned}$$

Here we use in $(*)$ that $\frac{d}{du}(\sqrt{u}X_i + \sqrt{1-u}Y_i) = \frac{1}{2\sqrt{u}}X_i - \frac{1}{2\sqrt{1-u}}Y_i$.

Claim I. *Condition on $\boldsymbol{Y}$. Then*

$$\sum_{i=1}^{n}\frac{1}{\sqrt{u}}\mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[X_i\cdot\frac{\partial f}{\partial x_i}\big(\boldsymbol{Z}(u)\big)\Big] = \sum_{i,j=1}^{n} M_{i,j}^X \mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{Z}(u))\Big]$$

[2]We use the following multivariate chain rule: let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}^n$ with $g(u) = (g_1(u), \ldots, g_n(u))$. Then $\frac{d}{du}(f(g(u)) = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(g(u))\cdot g_i'(u)$.

Proof of Claim. We consider $\boldsymbol{Y}$ as fixed for the remainder of the proof of this claim. Abbreviate $g_i(\boldsymbol{X}) := \frac{\partial f}{\partial x_i}\big(\sqrt{u}\boldsymbol{X} + \sqrt{1-u}\boldsymbol{Y}\big)$. Then

$$\begin{aligned}
\frac{1}{\sqrt{u}}\sum_{i=1}^{n}\mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[X_i\frac{\partial f}{\partial x_i}\big(\boldsymbol{Z}(u)\big)\Big] &= \frac{1}{\sqrt{u}}\sum_{i=1}^{n}\mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[X_i\cdot g_i(\boldsymbol{X})\Big] \\
&= \frac{1}{\sqrt{u}}\sum_{i=1}^{n}\sum_{j=1}^{n}M_{ij}^{X}\mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[\frac{\partial g_i}{\partial x_j}(\boldsymbol{X})\Big] \\
&= \frac{1}{\sqrt{u}}\sum_{i=1}^{n}\sum_{j=1}^{n}M_{ij}^{X}\mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[\frac{\partial^2 f}{\partial x_i\partial x_j}\big(\sqrt{u}\cdot\boldsymbol{X}+\sqrt{1-u}\cdot\boldsymbol{Y}\big)\cdot\sqrt{u}\Big] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}M_{ij}^{X}\mathop{\mathbb{E}}_{\boldsymbol{X}}\Big[\frac{\partial^2 f}{\partial x_i\partial x_j}\big(\boldsymbol{Z}(u)\big)\Big]
\end{aligned}$$

One can obtain an analogous expression for the Y_i-term (we skip the proof).
Claim II. *Condition on $\boldsymbol{X}$. Then*

$$\sum_{i=1}^{n}\frac{1}{\sqrt{1-u}}\mathop{\mathbb{E}}_{\boldsymbol{Y}}\Big[Y_i\cdot\frac{\partial f}{\partial x_i}\big(\boldsymbol{Z}(u)\big)\Big] = \sum_{i,j=1}^{n}M_{i,j}^{Y}\,\mathbb{E}\Big[\frac{\partial^2 f}{\partial x_i\partial x_j}(\boldsymbol{Z}(u))\Big]$$

□

9.2.2 Proof of Slepian's Lemma

After the previous Lemma we can give a natural condition when the increase is guaranteed to be non-negative. To get some intuition imagine we have a function $f(\boldsymbol{x}) = a_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i\neq j} c_{ij}x_i x_j$ which is a degree-2 polynomial. Then the expected function value is $\mathbb{E}_{\boldsymbol{X}\sim N(\boldsymbol{0},\boldsymbol{M})}[f(\boldsymbol{X})] = a_0 + \sum_{i\neq j} c_{ij}\,\mathbb{E}_{\boldsymbol{X}\sim N(\boldsymbol{0},\boldsymbol{M})}[X_i X_j]$ meaning that in particular any linear term is cancelled out. Then if $c_{ij} \geq 0$ and we increase $\mathbb{E}[X_i X_j] = M_{ij}$, then the expected function value would be increasing (or remain the same).

Lemma 9.12 (Functional form of Slepian's Lemma)**.** *Consider two Gaussian random vectors $\boldsymbol{X} \in \mathbb{R}^n$ and $\boldsymbol{Y} \in \mathbb{R}^n$ with*

$$\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2] \;\; \forall i \in [n] \quad \textit{and} \quad \mathbb{E}[|X_i - X_j|^2] \leq \mathbb{E}[|Y_i - Y_j|^2] \;\; \forall i,j \in [n]$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function with $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$ for all $i \neq j$. Then $\mathbb{E}[f(\boldsymbol{X})] \geq \mathbb{E}[f(\boldsymbol{Y})]$.

Proof. Using the formula from Lemma 9.11. we get

$$\frac{d}{du}\mathbb{E}[f(\boldsymbol{Z}(u))] = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{(M_{ij}^X - M_{ij}^Y)}_{\begin{cases} = 0 & \text{if } i = j, \\ \geq 0 & \text{if } i \neq j \end{cases}} \cdot \mathbb{E}\Big[\underbrace{\frac{\partial f^2}{\partial x_i \partial x_j}\big(\boldsymbol{Z}(u)\big)}_{\geq 0 \text{ if } i \neq j}\Big] \geq 0$$

As the expectation is increasing in u, we can conclude that $\mathbb{E}[f(\boldsymbol{Y})] = \mathbb{E}[f(\boldsymbol{Z}(0))] \leq \mathbb{E}[f(\boldsymbol{Z}(1))] = \mathbb{E}[f(\boldsymbol{X})]$. □

Now we can provide the proof of Slepian's Inequality:

Theorem 9.13 (Slepian's Inequality)**.** *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two finite Gaussian processes so that*

$$\mathbb{E}[X_t^2] = \mathbb{E}[Y_t^2] \quad \textit{and} \quad \mathbb{E}[(X_t - X_s)^2] \leq \mathbb{E}[(Y_t - Y_s)^2] \quad \forall s, t \in T$$

Then for any $\tau \in \mathbb{R}$ one has

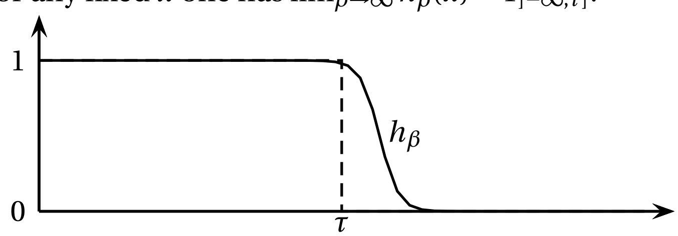
$$\Pr\Big[\sup_{t\in T} X_t \geq \tau\Big] \leq \Pr\Big[\sup_{t\in T} Y_t \geq \tau\Big].$$

Moreover $\mathbb{E}[\sup_{t\in T} X_t] \leq \mathbb{E}[\sup_{t\in T} Y_t]$.

Proof of Slepian's Inequality. We set $n := |T|$. For some parameter $\beta > 0$, let us define

$$h_\beta(x) := \frac{1}{1 + \exp(-\beta\cdot(\tau - x + \frac{1}{\sqrt{\beta}}))}$$

We make the observation that h_β is strictly decreasing in x, at least twice differentiable and for any fixed x one has $\lim_{\beta\to\infty} h_\beta(x) = 1_{]-\infty,\tau]}$.



We set $f_\beta(x) := \prod_{i=1}^n h_\beta(x_i)$ which is an approximation to the function $\mathbf{1}_{\max_{i=1,\dots,n}\{x_i\}\leq\tau}$. We can verify that for $i \neq j$ and $\boldsymbol{x} \in \mathbb{R}^n$ one has

$$\frac{\partial^2 f_\beta}{\partial x_i \partial x_j}(\boldsymbol{x}) = \underbrace{h'_\beta(x_i)}_{<0} \cdot \underbrace{h'_\beta(x_j)}_{<0} \prod_{k\in[n]\setminus\{i,j\}} \underbrace{h(x_k)}_{>0} > 0$$

Then by Lemma 9.12 one has

$$\Pr\Big[\max_{i=1,\dots,n} X_i \leq \tau\Big] = \lim_{\beta\to\infty} \mathbb{E}[f_\beta(\boldsymbol{X})] \geq \lim_{\beta\to\infty} \mathbb{E}[f_\beta(\boldsymbol{Y})] = \Pr\Big[\max_{i=1,\dots,n} Y_i \leq \tau\Big]$$

The "moreover" part follows from the integral identity. $\square$

9.2.3 The Sudakov-Fernique Inequality

We have stated earlier the geometric version of the Sudakov-Fernique Inequality (see Lemma 9.8) which only proves the "moreover" part of Slepian's lemma, but it allows that the variance of the random variables strictly increases. Now we give the probabilistic version and the proof:

Theorem 9.14 (Sudakov-Fernique Inequality)**.** *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two finite Gaussian processes so that*

$$\mathbb{E}[|X_t - X_s|^2] \leq \mathbb{E}[|Y_t - Y_s|^2] \quad \forall\, t \neq s$$

Then $\mathbb{E}[\sup_{t\in T} X_t] \leq \mathbb{E}[\sup_{t\in T} Y_t]$.

Proof. After reindexing we have $T = \{1,\dots,n\}$ and $X_i = \langle \boldsymbol{g}, \boldsymbol{x}_i\rangle$ and $Y_i = \langle \boldsymbol{g}, \boldsymbol{y}_i\rangle$ for some vectors $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ and $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$. Recall that the covariance matrix of $\boldsymbol{X}$ is $M^{\boldsymbol{X}}_{ij} = \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle = \frac{1}{2}(\|\boldsymbol{x}_i\|_2^2 + \|\boldsymbol{x}_j\|_2^2 - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2)$ and the assumption says that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 \leq \|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2^2$ for all $i, j \in T$. As before, we set $\boldsymbol{Z}(u) := \sqrt{u}\boldsymbol{X} + \sqrt{1-u}\boldsymbol{Y}$. For a parameter $\beta > 0$ we define a function

$$f_\beta(\boldsymbol{z}) := \frac{1}{\beta}\log\Big(\sum_{i=1}^n \exp(\beta z_i)\Big)$$

which is often called *soft maximum* function. This function is monotonically increasing and at least twice differentiable with $\lim_{\beta\to\infty} f_\beta(\boldsymbol{z}) = \max_{i=1,\dots,n} z_i$ for every fixed $\boldsymbol{z} \in \mathbb{R}^n$. It will be convinient to abbreviate coefficients $p_i(\boldsymbol{z}) := \frac{\exp(\beta z_i)}{\sum_{k=1}^n \exp(\beta z_k)}$. Note that $p_i(\boldsymbol{z}) \geq 0$ and $\sum_{i=1}^n p_i(\boldsymbol{z}) = 1$. One can easily verify that that 2nd derivatives of the softmax functions are

$$\frac{\partial^2 f_\beta(\boldsymbol{z})}{\partial^2 z_i} = \beta \cdot (p_i(\boldsymbol{z}) - p_i(\boldsymbol{z})^2) \quad \text{and} \quad \frac{\partial^2 f_\beta(\boldsymbol{z})}{\partial z_i \partial z_j} = -\beta \cdot p_i(\boldsymbol{z}) \cdot p_j(\boldsymbol{z})$$

We can simplify the contribution of $\boldsymbol{X}$ to the Gaussian Interpolation Formula as

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}^{\boldsymbol{X}} \cdot \frac{\partial f_\beta^2(\boldsymbol{z})}{\partial z_i \partial z_j} &= \beta \sum_{i=1}^{n} p_i \|\boldsymbol{x}_i\|_2^2 - \beta \sum_{i=1}^{n}\sum_{j=1}^{n} \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle \, p_i p_j \\
&= \beta \sum_{i=1}^{n} p_i \|\boldsymbol{x}_i\|_2^2 - \frac{\beta}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} (\|\boldsymbol{x}_i\|_2^2 + \|\boldsymbol{x}_j\|_2^2 - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2) p_i p_j \\
&= \underbrace{\beta \sum_{i=1}^{n} p_i \|\boldsymbol{x}_i\|_2^2 - 2 \cdot \frac{\beta}{2} \sum_{i=1}^{n} p_i \|\boldsymbol{x}_i\|_2^2 \underbrace{\sum_{j=1}^{n} p_j}_{=1}}_{=0} + \frac{\beta}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2
\end{aligned}$$

for any vector $\boldsymbol{z}$ with $p_i := p_i(\boldsymbol{z})$. Then using the *Gaussian Interpolation Formula* we obtain

$$\begin{aligned}
\frac{d}{du} \mathbb{E}[f_\beta(\boldsymbol{Z}(u))] &= \frac{1}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} (M_{ij}^X - M_{ij}^Y) \cdot \mathbb{E}\Big[\frac{\partial f_\beta^2}{\partial z_i \partial z_j}\big(\boldsymbol{Z}(u)\big)\Big] \\
&= \frac{\beta}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{\big(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 - \|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2^2\big)}_{\le 0} \cdot \mathbb{E}\Big[\underbrace{p_i(\boldsymbol{Z}(u))}_{\ge 0} \cdot \underbrace{p_j(\boldsymbol{Z}(u))}_{\ge 0}\Big] \le 0
\end{aligned}$$

As in a previous proof, this implies

$$\mathbb{E}\Big[\max_{i=1,\dots,n} X_i\Big] = \lim_{\beta\to\infty} \mathbb{E}\big[f_\beta(\underbrace{\boldsymbol{Z}(1)}_{=\boldsymbol{X}})\big] \le \lim_{\beta\to\infty} \mathbb{E}\big[f_\beta(\underbrace{\boldsymbol{Z}(0)}_{=\boldsymbol{Y}})\big] = \mathbb{E}\Big[\max_{i=1,\dots,n} Y_i\Big]$$

□

9.3 Sudakovs' Minoration Inequality

We can now give an answer to the problem of proving lower bounds on $\mathbb{E}[\sup_{t\in T} X_t]$ for non-independent Gaussian random variables. The beauty in the Sudakov-Fernique Comparsion Inequality is that we can now analyze $g(T)$ for non-orthogonal vectors by comparing it with $g(S)$ for orthogonal vectors S.

Lemma 9.15 (Sudakov's Inequality — Geometric version)**.** *Let $T = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ be vectors with $A \le \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \le B$ for $i \ne j$. Then*

$$\Theta\big(A\sqrt{\log(N)}\big) \le g(T) \le \Theta\big(B\sqrt{\log(N)}\big)$$

Lemma 9.16 (Sudakov's Inequality)**.** *Let $(X_t)_{t\in T}$ be a finite Gaussian process where $A \leq \mathbb{E}[|X_{t_1} - X_{t_2}|^2]^{1/2} \leq B$ for all $t_1 \neq t_2$. Then*

$$\Theta\Big(A\sqrt{\log(|T|)}\Big) \leq \mathbb{E}\Big[\sup_{t\in T} X_t\Big] \leq \Theta\Big(B\sqrt{\log(|T|)}\Big)$$

Proof. We only prove the lower bound; the upper bound is already proven in Lemma 9.7. Let $(Y_t)_{t\in T}$ be independent Gaussians with $\mathbb{E}[Y_t^2]^{1/2} = \frac{A}{\sqrt{2}}$. Then for $s \neq t$ one has $\mathbb{E}[|Y_t - Y_s|^2] = \mathbb{E}[Y_t^2] + \mathbb{E}[Y_s^2] = A^2 \leq \mathbb{E}[|X_t - X_s|^2]$. By Sudakov-Fernique Comparison Inequality (Lemma 9.14) and Lemma 9.7 for independent Gaussians we have

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big] \geq \mathbb{E}\Big[\sup_{t\in T} Y_t\Big] \geq \Omega\Big(A\sqrt{\log(|T|)}\Big)$$

$\square$

A useful quantity in the study of Gaussian processes is the following:

Definition 9.17. Let $\mathcal{X} = (X_t)_{t\in T}$ be a Gaussian process with distances $d(s,t) := \mathbb{E}[|X_s - X_t|^2]^{1/2}$ and $d(s,S) := \inf\{d(s,t) : t \in S\}$. For $\varepsilon > 0$ we define

$$N_\varepsilon(\mathcal{X}) := \min\big\{|S| : S \subseteq T \text{ and } \forall\, t \in T : d(t,S) \leq \varepsilon\big\}$$

If $\mathcal{X} = (X_{\boldsymbol{t}})_{\boldsymbol{t}\in T}$ happens to be a canonical Gaussian process with $T \subseteq \mathbb{R}^n$ then this definition corresponds to the covering number $N_\varepsilon(\mathcal{X}) = \bar{N}(T, \varepsilon B_2^n)$ (where centers need to be chosen from the set T). In that case we also use the notation $N_\varepsilon(T)$.

Theorem 9.18 (Sudakov's Minoration Inequality)**.** *Let $\mathcal{X} = (X_t)_{t\in T}$ be a Gaussian process. Then for any $\varepsilon > 0$ one has $\mathbb{E}[\sup_{t\in T} X_t] \geq \Omega(\varepsilon\sqrt{\log N_\varepsilon(\mathcal{X})})$.*

Proof. Pick a *maximal* set $S \subseteq T$ with $d(s,t) \geq \varepsilon$ for $s,t \in S$. Then $N_\varepsilon(\mathcal{X}) \leq |S|$ (as $d(t,S) \leq \varepsilon$ for all $t \in T$). Then applying the Lemma 9.16 to S we get

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big] \geq \mathbb{E}\Big[\sup_{t\in S} X_t\Big] \overset{\text{Lem 9.16}}{\geq} \Theta\big(\varepsilon\sqrt{\log(|S|)}\big) \geq \Theta\big(\varepsilon\sqrt{\log(N_\varepsilon(\mathcal{X}))}\big)$$

(note that if S can be picked with $|S| = \infty$, then $\mathbb{E}[\sup_{t\in T} X_t] = \infty$). $\square$

9.3.1 Application for Covering Numbers in $\mathbb{R}^n$

Now we can give a nice application to bounding the volume of a polytope with few vertices. Intuitively the claim says that a polytope contained in B_2^n needs exponentially many vertices in order to cover a significant fraction of the volume of B_2^n.

Theorem 9.19. *Let $P \subseteq B_2^n$ be a polytope with N vertices. Then*

$$\frac{Vol_n(P)}{Vol_n(B_2^n)} \leq \left(\frac{\log(N)}{n}\right)^{O(n)}$$

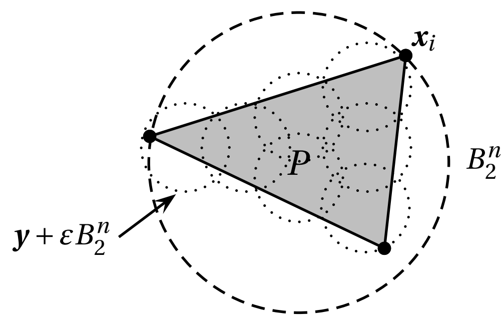
Proof. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be the vertices of P, that means $P = \text{conv}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$. Let $\varepsilon > 0$ be a parameter that we determine later. We use Sudakov's Minoration Inequality to obtain

$$\Theta(\varepsilon\sqrt{\log(N_\varepsilon(P))}) \overset{\text{Thm 9.18}}{\leq} g(P) \overset{(*)}{=} g(\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}) \overset{\|\boldsymbol{x}_i\|_2 \leq 1}{\leq} \Theta(\sqrt{\log(N)}) \quad (**)$$

Here we use in $(*)$ that a linear objective function over P is maximized at one of its vertices. Rearranging $(**)$ then tells us that we can cover P with at most

$$N_\varepsilon(P) \leq N^{\Theta(1/\varepsilon^2)}$$

many ε-radius balls.



Then considering a covering of P with ε-radius balls gives a volume bound of

$$\frac{\text{Vol}_n(P)}{\text{Vol}_n(B_2^n)} \leq N^{\Theta(1/\varepsilon^2)} \cdot \frac{\text{Vol}_n(\varepsilon B_2^n)}{\text{Vol}_n(B_2^n)} = N^{\Theta(1/\varepsilon^2)} \cdot \varepsilon^n \overset{\varepsilon := \sqrt{\log(N)/n}}{\leq} \left(\frac{\log(N)}{n}\right)^{Cn}$$

□

9.3.2 The Gaussian Contraction Principle

We want to state an inequality that is often useful. A function $\phi : \mathbb{R} \to \mathbb{R}$ is called a *contraction*, if $|\phi(x) - \phi(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$. For example the function $\phi(x) := |x|$ is a contraction.

Theorem 9.20 (Gaussian Contraction Inequality)**.** *Let $P \subseteq \mathbb{R}^n$ be a bounded symmetric set and let $\phi_1, \ldots, \phi_n : \mathbb{R} \to \mathbb{R}$ be contractions. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)} \left[\sup_{\boldsymbol{z} \in P} \left|\sum_{i=1}^n g_i \phi_i(z_i)\right|\right] \leq \mathop{\mathbb{E}}_{\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)} \left[\sup_{\boldsymbol{z} \in P} \left|\sum_{i=1}^n g_i z_i\right|\right]$$

This result can be easily proven by applying the Sudakov-Fernique Inequality in Theorem 9.14.

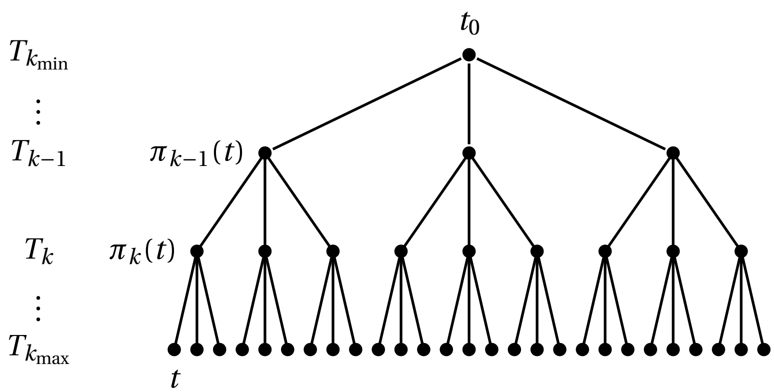
9.4 Dudley's Inequality

In this section we will answer the question whether one can provide a systematic upper bound on the supremum of a Gaussian process which is better than just taking the union bound as in Lemma 9.7. In some sense we will show that Sudakov's lower bound of $\Theta(\varepsilon\sqrt{\log N_\varepsilon(\mathcal{X})})$ is tight minus the issue that we need to sum over the whole range of possible ε's.

Theorem 9.21 (Discrete version of Dudley's Inequality)**.** *For any finite*[3] *Gaussian process $\mathcal{X} = (X_t)_{t\in T}$ one has*

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big] \le O(1)\cdot \sum_{\varepsilon\in 2^{\mathbb{Z}}} \varepsilon\cdot\sqrt{\log N_\varepsilon(\mathcal{X})}.$$

Proof. Let $T_k \subseteq T$ be minimal $(\frac{1}{2})^k$-nets of size $|T_k| = N_{(1/2)^k}(\mathcal{X})$. For every $t \in T$ we let $\pi_k(t) \in T_k$ denote the closest point, meaning that in particular $d(t,\pi_k(t)) \le (\frac{1}{2})^k$. We fix a $k_{\min} \in \mathbb{Z}$ and $k_{\max} \in \mathbb{Z}$ so that $|T_{k_{\min}}| = 1$ and $|T_{k_{\max}}| = T$ and write $T_{k_{\min}} = \{t_0\}$. This construction induces a tree with leaves T and root t_0 as depicted below:



Our goal is to find an upper bound on the quantity $\mathbb{E}[\sup_{t\in T}\{X_t - X_{t_0}\}]$ (which is the same as $\mathbb{E}[\sup_{t\in T} X_t]$ as $\mathbb{E}[X_{t_0}] = 0$). Then $\mathbb{E}[\sup_{t\in T}\{X_t - X_{t_0}\}]$ is upper

[3] In the available literature (see e.g. [Ver19, AAGM15]) this inequality is usually stated for any Gaussian process, not just finite ones. But that brings us to some subtle issues of how the expected supremum is actually defined. For example [AAGM15] defines $\mathbb{E}[\sup_{t\in T} X_t] := \sup\{\mathbb{E}[\sup_{t\in S} X_t] : S \subseteq T \text{ with } S \text{ finite}\}$. Then indeed the bound proven for finite subprocesses carries over.

bounded by the maximum length of a path $t = \pi_{k_{\max}}(t) \to \pi_{k_{\max}-1}(t) \to \ldots \to \pi_{k_{\min}+1}(t) \to \pi_{k_{\min}}(t) = t_0$ from a leaf to the root. To obtain Dudley's Inequality we will upper bound the expected maximum edge weight on *every level* which then naturally implies an upper bound on the expected maximum weight of every t_0-t path.

We first bound the expected maximum stretch on one level:

Claim. *For every k one has* $\mathbb{E}[\sup_{t\in T}\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\}] \leq O(1)\cdot(\frac{1}{2})^k\sqrt{\log|T_k|}$.

Proof of Claim. For a fixed $t \in T$ we have

$$d(\pi_k(t), \pi_{k-1}(t)) \leq d(\pi_k(t), t) + d(\pi_{k-1}(t), t) \leq \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{k-1} \leq 4\cdot\left(\frac{1}{2}\right)^k.$$

The number of *different* pairs $(\pi_k(t), \pi_{k-1}(t))$ is bounded by $|T_k|\cdot|T_{k-1}| \leq |T_k|^2$. Then using the upper bound from Lemma 9.16 we obtain

$$\mathbb{E}\Big[\sup_{t\in T}\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\}\Big] \leq O(1)\cdot 4\cdot\left(\frac{1}{2}\right)^k\sqrt{\log(|T_k|^2)}$$

which is of the claimed form. □

Now we can finish the argument (using that $\mathbb{E}[X_{t_0}] = 0$):

$$\begin{aligned}
\mathbb{E}\Big[\sup_{t\in T}\{X_t - X_{t_0}\}\Big] &= \mathbb{E}\Big[\sup_{t\in T}\Big\{\sum_{k=k_{\min}+1}^{k_{\max}} X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\Big\}\Big] \\
&\leq \sum_{k=k_{\min}+1}^{k_{\max}} \mathbb{E}\Big[\sup_{t\in T}\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\}\Big] \\
&\leq O(1)\cdot\sum_{k=k_{\min}+1}^{k_{\max}} (1/2)^k\sqrt{\log(N_{(1/2)^k}(\mathcal{X}))}
\end{aligned}$$

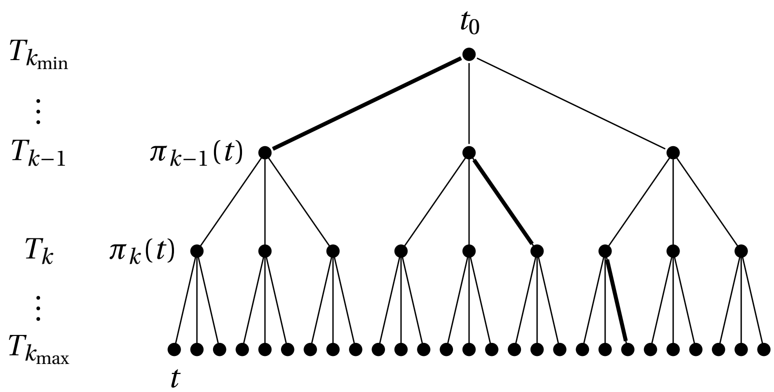
which is the desired bound. □

Theorem 9.21 can also be restated in the form $\mathbb{E}[\sup_{t\in T} X_t] \leq O(1)\cdot\int_0^\infty \sqrt{\log N_\varepsilon(\mathcal{X})}\,d\varepsilon$ which is called *Dudley's Integral Inequality*.

9.5 Generic Chaining and Talagrand's Majorizing Measure Theorem

While for many settings, Dudley's chaining argument gives already a tight bound, we will see a construction in the exercises where the gap between Dudley's bound and the real expected supremum is unbounded. To get some intuition where

Dudley's argument is potentially loose, consider a random process $\mathcal{X} = (X_t)_{t\in T}$ and the tree that we have used to prove Theorem 9.21. For Dudley's bound we have upper bounded the maximum weight of edges separatedly on every level. But it is not hard to see that in general the combination of those maximum weight edges do not form a path from a leaf to the root!



We will now discuss a refined bound due to Talagrand. Again we use the notation $d(t,S) := \inf_{s\in S} d(t,s)$ for a subset $S \subseteq T$.

Definition 9.22. Let T be a set with metric d. We call a sequence $(T_k)_{k=0,\dots,\infty}$ with $T_k \subseteq T$ an *admissible sequence* if

$$|T_0| = 1 \quad \text{and} \quad |T_k| \le 2^{2^k} \quad \forall k \in \{1,2,\dots\}$$

Then

$$\gamma_2(T,d) := \inf_{(T_k)_k \text{ admissible}} \left\{ \sup_{t\in T} \sum_{k=0}^{\infty} 2^{k/2} d(t,T_k) \right\}$$

Intuitively the bound says that we can select a "net" T_k and we need to pay for the maximum sum of distances of any t to all the nets, weighting the distance to T_k by $2^{k/2}$. If $\mathcal{X} = (X_t)_{t\in T}$ is a Gaussian process then we write $\gamma_2(\mathcal{X}) := \gamma_2(T,d)$ where $d(s,t) := \mathbb{E}[|X_s - X_t|^2]^{1/2}$ is the usual metric. We can now prove that the quantity $\gamma_2(\mathcal{X})$ is an upper bound on the expected supremum:

Theorem 9.23. *Let $\mathcal{X} = (X_t)_{t\in T}$ be a Gaussian process. Then*

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \le O(1)\cdot\gamma_2(\mathcal{X}).$$

Proof. Fix an admissible sequence $(T_k)_{k=0,\dots,\infty}$ with $T_0 = \{t_0\}$ and let $\pi_k(t) \in T_k$ denote the closest point to t.

Claim I. For $u \ge 10$, one has

$$(*) := \Pr\left[\sup_{t\in T} \left|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right| \le Cu2^{k/2}\cdot d\big(\pi_k(t),\pi_{k-1}(t)\big)\ \forall k\in\mathbb{N}\right] \ge 1-2\exp(-u^2)$$

Proof of Claim I. For a fixed k and t we have by Lemma 9.4.(C) that

$$\Pr\Big[|X_{\pi_k(t)} - X_{\pi_{k-1}(t-1)}| > Cu2^{k/2} \cdot d\big(\pi_k(t), \pi_{k-1}(t)\big)\Big] \leq 2\exp(-8u^2 2^k)$$

for some constant $C > 0$. Then we can bound

$$\begin{aligned}
(*) \quad &\overset{\text{union bound}}{\leq} \sum_{k=1}^{\infty} \sum_{s_1 \in T_k, s_2 \in T_{k-1}} \Pr\big[|X_{s_1} - X_{s_2}| > Cu2^{k/2} \cdot d(s_1, s_2)\big] \\
&\leq \sum_{k=1}^{\infty} \underbrace{|T_k| \cdot |T_{k-1}|}_{\leq 2^{2^{k+1}}} \cdot 2\exp\big(-8u^2 2^k\big) \leq 2\exp(-u^2)
\end{aligned}$$

if, say $u \geq 10$. So the claim is proven. $\square$

Now fix a value of u and suppose the event in Claim I happens. Then

$$\begin{aligned}
\sup_{t \in T} |X_t - X_{t_0}| &= \sup_{t \in T} \Big| \sum_{k=1}^{\infty} (X_{\pi_k(t)} - X_{\pi_{k-1}(t)}) \Big| \leq \sup_{t \in T} \sum_{k=1}^{\infty} \big| X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \big| \\
&\leq \sup_{t \in T} \sum_{k=1}^{\infty} Cu2^{k/2} \cdot d\big(\pi_k(t), \pi_{k-1}(t)\big) \leq Cu \cdot \gamma_2(T)
\end{aligned}$$

If we think of u as the random variable that gives the smallest possible value that makes Claim I true, then clearly $\mathbb{E}[u] \leq O(1)$ and the claim follows. $\square$

The amazing result (again due to Talagrand) is that $\gamma_2(\mathcal{X})$ is always a constant factor approximation to the real expected supremum for any Gaussian process.

Theorem 9.24 (Talagrand's Majorizing Measure Theorem [Tal87]). *Let $\mathcal{X} = (X_t)_{t \in T}$ be a Gaussian process. Then*

$$C_1 \gamma_2(\mathcal{X}) \leq \mathbb{E}\Big[\sup_{t \in T} X_t\Big] \leq C_2 \gamma_2(\mathcal{X})$$

for universal constants $C_1, C_2 > 0$.

The proof of the lower bound is rather involved and we will have to omit it here.

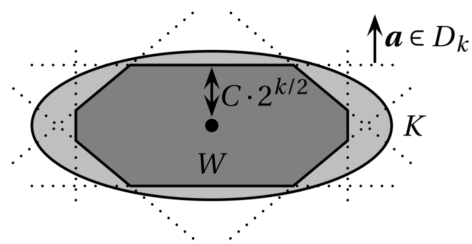
9.5.1 Approximating symmetric convex bodies

We want to describe a geometric implication of Talagrand's Majorizing Measure Theorem that can be tremendously useful. We can prove that for every symmetric convex body K there is an included body $W \subseteq K$ that has only few constraints that are close to the origin.

Theorem 9.25. *For any symmetric convex body $K \subseteq \mathbb{R}^n$ with $\mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_K] = 1$, there is a symmetric convex body*

$$W := \left\{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{d}, \boldsymbol{x}\rangle| \leq C \cdot 2^{k/2}\ \forall k \in \mathbb{N}\ \forall \boldsymbol{d} \in D_k\right\} \subseteq K$$

where $D_k \subseteq S^{n-1}$ with $|D_k| \leq 2^{2^{k+1}}$ and $C > 0$ is a universal constant.



Proof. We write $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x}\rangle \leq 1\ \forall \boldsymbol{a} \in T\}$ with $\boldsymbol{0} \in T$ (one can choose for example $T := K^\circ$, but our argument does not need convexity, so in case K is a polytope, T can be chosen as the facet normal vectors plus the origin). Consider the Gaussian process $\{X_{\boldsymbol{a}}\}_{\boldsymbol{a}\in T}$ with variables $X_{\boldsymbol{a}} := (\langle \boldsymbol{a}, \boldsymbol{g}\rangle)$ where $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$. The distance metric of that Gaussian process is simply $d(\boldsymbol{a}, \boldsymbol{b}) = \mathbb{E}[|X_{\boldsymbol{a}} - X_{\boldsymbol{b}}|^2]^{1/2} = \|\boldsymbol{a} - \boldsymbol{b}\|_2$. Note that $\|\boldsymbol{x}\|_K = \mathbb{E}[\sup_{\boldsymbol{a}\in T}\langle \boldsymbol{a}, \boldsymbol{x}\rangle]$. We apply *Talagrands Majorizing Theorem* (Theorem 9.24) and obtain an *admissible sequence* $\{T_k\}_{k=0,\ldots,\infty}$ with

$$(i)\ \ T_k \subseteq T \text{ for all } k \in \mathbb{Z}_{\geq 0}, \quad (ii)\ \ T_0 = \{\boldsymbol{0}\}, \quad (iii)\ \ |T_k| \leq 2^{2^k} \text{ for } k \in \mathbb{N}$$

so that

$$\sup_{\boldsymbol{a}\in T}\left\{\sum_{k=0}^{\infty} 2^{k/2} \cdot d(\boldsymbol{a}, T_k)\right\} \overset{\text{Thm 9.24}}{\leq} O(1) \cdot \mathbb{E}\left[\sup_{\boldsymbol{a}\in T} X_{\boldsymbol{a}}\right] \overset{\text{assumption}}{\leq} O(1)$$

For each $\boldsymbol{a} \in T$, let $\pi_k(\boldsymbol{a}) \in T_k$ be the closest element in T_k, i.e. $d(\boldsymbol{a}, \pi_k(\boldsymbol{a})) = d(\boldsymbol{a}, T_k)$. Now we have everything in place to define the body W that approximates K. We set

$$W := \left\{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{x}\rangle| \leq C \cdot 2^{k/2}\|\boldsymbol{a} - \boldsymbol{b}\|_2 \quad \forall k \in \mathbb{N}\ \forall \boldsymbol{a} \in T_k\ \forall \boldsymbol{b} \in T_{k-1}\right\}$$

where we will choose $C > 0$ small enough. We want to prove that $W \subseteq K$. First note that for any $\boldsymbol{a} \in T$ we can write

$$\boldsymbol{a} = \sum_{k=1}^{\infty}(\pi_k(\boldsymbol{a}) - \pi_{k-1}(\boldsymbol{a})).$$

Now fix any $\boldsymbol{x} \in W$ and $\boldsymbol{a} \in T$. Then

$$\begin{aligned}
|\langle \boldsymbol{a}, \boldsymbol{x}\rangle| &\leq \sum_{k=1}^{\infty} |\langle \pi_k(\boldsymbol{a}) - \pi_{k-1}(\boldsymbol{a}), \boldsymbol{x}\rangle| \\
&\stackrel{\boldsymbol{x}\in W}{\leq} \sum_{k=1}^{\infty} C \cdot 2^{k/2} \cdot \underbrace{d(\pi_k(\boldsymbol{a}), \pi_{k-1}(\boldsymbol{a}))}_{\leq d(\boldsymbol{a},T_k)+d(\boldsymbol{a},T_{k-1})} \leq O(C) \cdot \underbrace{\sum_{k=0}^{\infty} 2^{k/2} d(\boldsymbol{a}, T_k)}_{\leq O(1)} \leq O(C) \stackrel{C \text{ small}}{\leq} 1
\end{aligned}$$

and hence $\boldsymbol{x} \in K$. It remains to bring the constraints of W into the claimed format. We set $D_k := \{\frac{\boldsymbol{a}-\boldsymbol{b}}{\|\boldsymbol{a}-\boldsymbol{b}\|_2} : \boldsymbol{a} \in T_k, \boldsymbol{b} \in T_{k-1}\}$ for $k \in \mathbb{N}$. Then $|D_k| \leq |T_k| \cdot |T_{k-1}| \leq 2^{2^k} \cdot 2^{2^{k-1}} \leq 2^{2^{k+1}}$ and the claim is proven. □

By arranging the sets $T_0, T_1, T_2, \ldots$ as one sequence of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \ldots$ Theorem 9.25 can be conviniently rephrased as follows:

Corollary 9.26. *For any symmetric convex body $K \subseteq \mathbb{R}^n$ there is a sequence $\{\boldsymbol{a}_\ell\}_{\ell\in\mathbb{N}} \subseteq S^{n-1}$ of unit vectors so that*

$$W := \left\{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{a}_\ell, \boldsymbol{x}\rangle| \leq \beta_\ell \ \forall \ell \in \mathbb{N}\right\} \quad \textit{with} \quad \beta_\ell := \frac{C\sqrt{\log(2\ell)}}{\mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_K]}$$

satisfies $W \subseteq K$. Here, $C > 0$ is a universal constant.

We want to emphasize that this result in Cor 9.26 is *so* tight that the body satisfies $\mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_W] \leq O(1) \cdot \mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_K]$. Even stronger, simply by applying a union bound over all the constraints defining W, one can derive that $\mathbb{E}[\|\boldsymbol{X}\|_W] \leq O(1) \cdot \mathbb{E}_{\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_K]$ for any subgaussian random vector $\boldsymbol{X}$.

9.6 Subgaussian random variables and Talagrand's Comparison Inequality

The reader should observe that many of the upper bounds on $\mathbb{E}[\sup_{t\in T} X_t]$ which we have seen so far, only use the *tail bounds* that we know for Gaussian random variables X_t. This applies for example to the upper bound in Lemma 9.7, Dudley's Inequality (Theorem 9.21) or Theorem 9.23. In turn this means that the same bounds would apply to any "non-Gaussian" random variables as long as these satisfy Gaussian tail bounds. At this point we recommend the reader to review the *subgaussian norm* introduced in Section 3.4.

For example we can use Dudley's bound or Talagrand's bound (Theorem 9.23) to upper bound $\mathbb{E}[\sup_{t\in T} X_t]$ in terms of the $\|\cdot\|_{\psi_2}$-norms. But often this is still tedious! Amazingly we can simply upper bound $\mathbb{E}[\sup_{t\in T} X_t]$ by the value of a *dominating Gaussian process.*

Theorem 9.27 (Talagrand's Comparison Inequality). *Let $(X_t)_{t\in T}$ be a mean-zero random process and let $(Y_t)_{t\in T}$ be a Gaussian random process. Assume that*

$$\|X_s - X_t\|_{\psi_2} \leq \mathbb{E}[|Y_s - Y_t|^2]^{1/2} \quad \forall s,t \in T$$

Then

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big] \leq C \cdot \mathbb{E}\Big[\sup_{t\in T} Y_t\Big]$$

for a universal constant $C > 0$.

Proof. We abbreviate $d_X(s,t) := \|X_s - X_t\|_{\psi_2}$ and $d_Y(s,t) := \mathbb{E}[|Y_s - Y_t|^2]^{1/2}$ for $s,t \in T$. First we make the observation that the proof in Theorem 9.23 only uses tail bounds for Gaussians and applies as well to the random process $(X_t)_{t\in T}$, meaning that $\mathbb{E}[\sup_{t\in T} X_t] \leq \gamma_2(T, d_X)$. Next, the assumption tells us that $d_X(s,t) \leq d_Y(s,t)$ for all $s,t \in T$ and so by monotonicity we have $\gamma_2(T,d_X) \leq \gamma_2(T,d_Y)$. Finally *Talagrand's Majorizing Measure Theorem* [Tal87] shows that up to a constant factor, $\gamma_2(T,d_Y)$ is also a lower bound on $\mathbb{E}[\sup_{t\in T} Y_t]$. We summarize this to

$$\mathbb{E}\Big[\sup_{t\in T} X_t\Big] \overset{\text{Theorem 9.23}}{\leq} C_1 \cdot \gamma_2(T,d_X) \overset{\text{monotonicity}}{\leq} C_1 \cdot \gamma_2(T,d_Y) \leq C_1 C_2 \cdot \mathbb{E}\Big[\sup_{t\in T} Y_t\Big]$$

for some constants $C_1, C_2 > 0$. □

We would like to point out that this is an incredibly powerful principle. While we will only see two such application of Talagrand's Comparison Inequality here, the book of Vershynin [Ver19] has several more.

9.6.1 The inequality of Maurey and Pisier

We prove a variant of an inequality due to Maurey and Pisier [MP76].

Theorem 9.28 (Maurey-Pisier). *There is a universal constant $C > 0$ so that the following holds: For any symmetric convex body $K \subseteq \mathbb{R}^n$ one has*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}[\|\boldsymbol{x}\|_K] \leq C \cdot \mathop{\mathbb{E}}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{y}\|_K].$$

Proof. We note that for any vector $\boldsymbol{b} \in \mathbb{R}^n$ and some constant $C_0 > 0$ one has $\|\langle \boldsymbol{b}, \boldsymbol{x}\rangle\|_{\psi_2} \leq C_0 \|\boldsymbol{b}\|_2$ (where $\boldsymbol{x} \sim \{-1,1\}^n$) as one can easily derive from Lemma 3.17.

We can write $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle \leq 1 \;\forall i \in T\}$ for some index set T. Consider the mean zero random processes $(X_i)_{i\in T}$ with $X_i := \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle$ and $\boldsymbol{x} \sim \{-1,1\}^n$ as well as the Gaussian process $(Y_i)_{i\in T}$ with $Y_i := \langle C_0\boldsymbol{a}_i, \boldsymbol{y}\rangle$ and $\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$. Then for indices $i, j \in T$ we have

$$\|X_i - X_j\|_{\psi_2} = \|\langle \boldsymbol{a}_i - \boldsymbol{a}_j, \boldsymbol{x}\rangle\|_{\psi_2} \leq C_0\|\boldsymbol{a}_i - \boldsymbol{a}_j\|_2 = \mathbb{E}[|Y_i - Y_j|^2]^{1/2}$$

Then by Talagrand's Comparison Inequality (Theorem 9.27) we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}[\|\boldsymbol{x}\|_K] = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}\Big[\sup_{i\in T}\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle\Big] &\leq C_1 \mathop{\mathbb{E}}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}\Big[\sup_{i\in T}\langle C_0\boldsymbol{a}_i, \boldsymbol{y}\rangle\Big] \\ &= C_0C_1 \mathop{\mathbb{E}}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{y}\|_K] \end{aligned}$$

□

9.7 Concentration for Gaussian Random Matrices

The goal for remainder of this chapter is to prove the tight bound of Dvoretzky's Theorem. The key tool is a very tight — and very flexible — deviation inequality for random matrices. We write $N^{m\times n}(0,1)$ as the distribution over *Gaussian random matrices* $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ where all entries are drawn as $A_{ij} \sim N(0,1)$ independently. Recall that a set T is symmetric if $-T = T$.

Theorem 9.29. *Let $K \subseteq \mathbb{R}^m$ be a symmetric convex body with $\frac{1}{b}B_2^m \subseteq K$ and let $\boldsymbol{A} \sim N^{m\times n}(0,1)$ be Gaussian random matrix. For any symmetric set $T \subseteq \mathbb{R}^n$ one has*

$$\mathbb{E}\Big[\sup_{\boldsymbol{x}\in T}\Big|\|\boldsymbol{A}\boldsymbol{x}\|_K - \mathbb{E}\big[\|\boldsymbol{A}\boldsymbol{x}\|_K\big]\Big|\Big] \leq O(b)\cdot g(T)$$

Proof. By scaling K and T we may assume that $b = 1$ and so $B_2^m \subseteq K$. For $\boldsymbol{x} \in \mathbb{R}^n$ we define the random variable

$$X_{\boldsymbol{x}} := \|\boldsymbol{A}\boldsymbol{x}\|_K - \mathbb{E}[\|\boldsymbol{A}\boldsymbol{x}\|_K]$$

The claim is that $\mathbb{E}[\sup_{\boldsymbol{x}\in T}|X_{\boldsymbol{x}}|] \leq O(g(T))$. We note that $(X_{\boldsymbol{x}})_{\boldsymbol{x}\in T}$ is *not* a Gaussian process, but at least $\mathbb{E}[X_{\boldsymbol{x}}] = 0$ for all $\boldsymbol{x}$. Moreover, we should explain why each random variable $X_{\boldsymbol{x}}$ is sub-gaussian. As $B_2^m \subseteq K$ we know that the map $\boldsymbol{y} \mapsto \|\boldsymbol{y}\|_K$ is 1-Lipschitz and for a fixed vector $\boldsymbol{x} \in \mathbb{R}^n$ we know that $\boldsymbol{A}\boldsymbol{x} \sim \|\boldsymbol{x}\|_2 N(\boldsymbol{0}, \boldsymbol{I}_m)$. Then by Theorem 3.10 this means that in our recently developed notation one has $\big\|\|\boldsymbol{A}\boldsymbol{x}\|_K\big\|_{\psi_2} \leq C\|\boldsymbol{x}\|_2$ for some constant $C > 0$.

Now back to the actual proof. We will compare the mean-zero random process $(X_{\boldsymbol{x}})_{\boldsymbol{x}\in T}$ and with the canonical Gaussian process $(Y_{\boldsymbol{x}})_{\boldsymbol{x}\in T}$ defined by $Y_{\boldsymbol{x}} :=$

$C \cdot \langle \boldsymbol{g}, \boldsymbol{x} \rangle$ where $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$ and C is a large enough constant. Assuming that we can indeed prove that

$$\|X_{\boldsymbol{x}} - X_{\boldsymbol{y}}\|_{\psi_2} \overset{\text{to prove!}}{\leq} C\|\boldsymbol{x}-\boldsymbol{y}\|_2 = \mathbb{E}\left[|Y_{\boldsymbol{x}} - Y_{\boldsymbol{y}}|^2\right]^{1/2} \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

then by *Talagrand's Comparison Inequality* (Theorem 9.27) we can conclude that[4]

$$\mathbb{E}\left[\sup_{\boldsymbol{x}\in T} |X_{\boldsymbol{x}}|\right] \overset{\text{Talagrand}}{\leq} O(1) \cdot \mathbb{E}\left[\sup_{\boldsymbol{x}\in T} |Y_{\boldsymbol{x}}|\right] = O(1) \cdot \mathop{\mathbb{E}}_{\boldsymbol{g}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}\left[\sup_{\boldsymbol{x}\in T} |\langle \boldsymbol{g}, \boldsymbol{x}\rangle|\right] \overset{T \text{ symmetric}}{=} O(g(T)).$$

The first step is to analyze the subgaussian norm of individual random variables more carefully:

Claim I. *Fix $\boldsymbol{a} \in \mathbb{R}^m$, $s \geq 0$ and consider the random variable $f(\boldsymbol{b}) := \|\boldsymbol{a} + s\boldsymbol{b}\|_K$ where $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{I}_m)$. Then $\|f(\boldsymbol{b}) - \mathbb{E}[f(\boldsymbol{b})]\|_{\psi_2} \leq O(s)$.*

Proof of Claim I. By Theorem 3.10 it suffices to prove that the function f is s-Lipschitz. And indeed for $\boldsymbol{b}, \boldsymbol{b}' \in \mathbb{R}^m$ one has

$$|f(\boldsymbol{b}) - f(\boldsymbol{b}')| = \big|\|\boldsymbol{a} + s\boldsymbol{b}\|_K - \|\boldsymbol{a} + s\boldsymbol{b}'\|_K\big| \leq s\|\boldsymbol{b} - \boldsymbol{b}'\|_K \overset{B_2^m \subseteq K}{\leq} s\|\boldsymbol{b} - \boldsymbol{b}'\|_2$$

Next, we show the required claim for unit length vectors.

Claim II. *For $\boldsymbol{x}, \boldsymbol{y} \in S^{n-1}$ one has $\|X_{\boldsymbol{x}} - X_{\boldsymbol{y}}\|_{\psi_2} \leq C\|\boldsymbol{x} - \boldsymbol{y}\|_2$ for a large enough constant $C > 0$.*

Proof of claim. Define $\boldsymbol{u} := \frac{\boldsymbol{x}+\boldsymbol{y}}{2}$ and $\boldsymbol{v} := \frac{\boldsymbol{x}-\boldsymbol{y}}{2}$. We keep in mind that $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$, $\boldsymbol{y} = \boldsymbol{u} - \boldsymbol{v}$. Note that $\boldsymbol{u} \perp \boldsymbol{v}$ and so $\boldsymbol{A}\boldsymbol{u} \sim \|\boldsymbol{u}\|_2 N(\boldsymbol{0}, \boldsymbol{I}_m)$ and $\boldsymbol{A}\boldsymbol{v} \sim \|\boldsymbol{v}\|_2 N(\boldsymbol{0}, \boldsymbol{I}_m)$ are independent random vectors. In the following we write we write $\|\cdot\|_{\psi_2(\boldsymbol{A}\boldsymbol{u})}$ for the subgaussian norm of a random variable where we have conditioned on the outcome of[5] $\boldsymbol{A}\boldsymbol{u}$. Then conditioning on any outcome of $\boldsymbol{A}\boldsymbol{u}$ we have

$$\begin{aligned}
\Big\| \|\boldsymbol{A}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{v}\|_K - \mathbb{E}[\|\boldsymbol{A}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{v}\|_K \mid \boldsymbol{A}\boldsymbol{u}] \Big\|_{\psi_2(\boldsymbol{A}\boldsymbol{u})} &\overset{\substack{\boldsymbol{a}:=\boldsymbol{A}\boldsymbol{u},\\ s:=\|\boldsymbol{v}\|_2,\\ \boldsymbol{b}\sim N(\boldsymbol{0},\boldsymbol{I}_m)}}{=} \Big\| \|\boldsymbol{a} + s\boldsymbol{b}\|_k - \mathbb{E}[\|\boldsymbol{a} + s\boldsymbol{b}\|_K] \Big\|_{\psi_2} \\
&\overset{\text{Claim I}}{\leq} O(\|\boldsymbol{v}\|_2)
\end{aligned}$$

Analogously we have

$$\Big\| \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{A}\boldsymbol{v}\|_K - \mathbb{E}[\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{A}\boldsymbol{v}\|_K \mid \boldsymbol{A}\boldsymbol{u}] \Big\|_{\psi_2(\boldsymbol{A}\boldsymbol{u})} \leq O(\|\boldsymbol{v}\|_2)$$

[4] Here we have the small technicality that we have $|\cdot|$ on the left hand side. That means we are actually interested in the expected supremum of the random process $(\pm X_{\boldsymbol{x}})_{\boldsymbol{x}\in T}$. Using Talagrand we can relate this to the expected supremum of the Gaussian process $(\pm Y_{\boldsymbol{x}})_{\boldsymbol{x}\in T}$ which is the same as the expected supremum of $(Y_{\boldsymbol{x}})_{\boldsymbol{x}\in T}$ because we assumed that T is symmetric.

[5] Which does *not* mean that we have conditioned on the outcome of all of $\boldsymbol{A}$.

In fact the random variables $\boldsymbol{Au}+\boldsymbol{Av}$ and $\boldsymbol{Au}-\boldsymbol{Av}$ have identical distributions (regardless whether we condition on $\boldsymbol{Au}$) or not. Then using the triangle inequality (Lemma 3.17.(iii)) one has

$$\begin{aligned}
& \Big\| \|\boldsymbol{Au}+\boldsymbol{Av}\|_K - \|\boldsymbol{Au}-\boldsymbol{Av}\|_K \Big\|_{\psi_2(\boldsymbol{Au})} \\
= \; & \Big\| \big(\|\boldsymbol{Au}+\boldsymbol{Av}\|_K - \mathbb{E}[\|\boldsymbol{Au}+\boldsymbol{Av}\|_K \mid \boldsymbol{Au}]\big) - \big(\|\boldsymbol{Au}-\boldsymbol{Av}\|_K - \mathbb{E}[\|\boldsymbol{Au}-\boldsymbol{Av}\|_K \mid \boldsymbol{Au}]\big) \Big\|_{\psi_2(\boldsymbol{Au})} \\
\overset{\text{Lem. 3.17.(iii)}}{\leq} \; & \Big\| \|\boldsymbol{Au}+\boldsymbol{Av}\|_K - \mathbb{E}[\|\boldsymbol{Au}+\boldsymbol{Av}\|_K \mid \boldsymbol{Au}] \Big\|_{\psi_2(\boldsymbol{Au})} + \Big\| \|\boldsymbol{Au}-\boldsymbol{Av}\|_K - \mathbb{E}[\|\boldsymbol{Au}-\boldsymbol{Av}\|_K \mid \boldsymbol{Au}] \Big\|_{\psi_2(\boldsymbol{Au})} \\
\leq \; & O(\|\boldsymbol{v}\|_2)
\end{aligned}$$

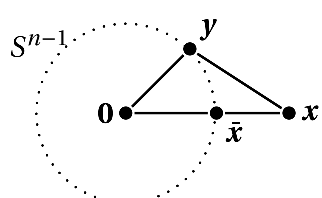
Since this is true for every conditioning $\boldsymbol{Au}$ it must also hold unconditioned (one can derive this easily from the definition of ψ_2). Hence

$$\Big\| \|\boldsymbol{Ax}\|_K - \|\boldsymbol{Ay}\|_K \Big\|_{\psi_2} = \Big\| \|\boldsymbol{Au}+\boldsymbol{Av}\|_K - \|\boldsymbol{Au}-\boldsymbol{Av}\|_K \Big\|_{\psi_2} \leq O(\|\boldsymbol{v}\|_2) = O(\|\boldsymbol{x}-\boldsymbol{y}\|_2)$$

which finishes Claim II. □

Now we consider the case for arbitrary vectors $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$. By symmetry we may assume $\|\boldsymbol{x}\|_2 \geq \|\boldsymbol{y}\|_2$. Also we have already argued that $\|X_{\boldsymbol{x}}\|_{\psi_2} \leq O(\|\boldsymbol{x}\|_2)$ which covers the case $\|\boldsymbol{y}\|_2 = 0$. Then after scaling we may assume that $\|\boldsymbol{x}\|_2 \geq 1$ and $\|\boldsymbol{y}\|_2 = 1$. We write $\boldsymbol{x} = s\bar{\boldsymbol{x}}$ with $s = \|\boldsymbol{x}\|_2$ and $\|\bar{\boldsymbol{x}}\|_2 = 1$. Then $X_{\boldsymbol{x}} = s \cdot X_{\bar{\boldsymbol{x}}}$ and so

$$\begin{aligned}
\|X_{\boldsymbol{x}} - X_{\boldsymbol{y}}\|_{\psi_2} &= \|(s-1)X_{\bar{\boldsymbol{x}}} + (X_{\bar{\boldsymbol{x}}} - X_{\boldsymbol{y}})\|_{\psi_2} \\
&\overset{\text{Lem. 3.17}}{\leq} (s-1)\|X_{\boldsymbol{x}}\|_{\psi_2} + \|X_{\bar{\boldsymbol{x}}} - X_{\boldsymbol{y}}\|_{\psi_2} \\
&\leq O(1) \cdot \Big((s-1) + \|\bar{\boldsymbol{x}} - \boldsymbol{y}\|_2 \Big) \\
&= O(1) \cdot \Big(\underbrace{\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2}_{\leq \|\boldsymbol{x}-\boldsymbol{y}\|_2} + \underbrace{\|\bar{\boldsymbol{x}} - \boldsymbol{y}\|_2}_{\leq \|\boldsymbol{x}-\boldsymbol{y}\|_2} \Big) \\
&\leq O(1) \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_2
\end{aligned}$$



That finishes the proof. □

We would like to point out that this theorem can be proven not just for norms but also for the more general class of *positive-homogeneous, subadditive* functions. Again, see [Ver19] for details.

9.8 The tight version of Dvoretzky's Theorem

Recall that in Chapter 5 we have proven the following version fo Dvoretzky's Theorem:

Theorem (Theorem 5.6)**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $0 < \varepsilon \leq \frac{1}{2}$. Then there exists a subspace V of dimension $k := \Theta(\frac{\varepsilon^2}{\log(1/\varepsilon)} \log(n))$ so that $K \cap V$ is $(1+\varepsilon)$-spherical.*

It turns out that the dependence on ε is not tight and with some work one can remove the $\log(1/\varepsilon)$-term. Now that we are closing in on the proof of Dvoretzky's Theorem we want to discuss how we actually generate a random k-dimensional subspace $F \subseteq \mathbb{R}^n$. In Chapter 5 we have typically selected a matrix $\boldsymbol{U} \in \mathbb{R}^{n\times k}$ so that the columns are orthonormal. This had the benefit that $F \cap S^{n-1} = \{\boldsymbol{U}\boldsymbol{x} : \boldsymbol{x} \in S^{k-1}\}$ are precisely the unit vectors in that subspace. On the other hand, the machinery that we developed here only applies to random Gaussians. Hence we want to draw a Gaussian random matrix $\boldsymbol{A} \sim N^{n\times k}(0,1)$ and use the subspace $F := \text{span}\{\boldsymbol{A}\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^k\}$ spanned by the columns of $\boldsymbol{A}$. Note that by rotational symmetry this is again a *uniform* k-dimensional subspace. Consider the map $T : \mathbb{R}^k \to \mathbb{R}^n$ with $T(\boldsymbol{x}) := \frac{1}{a_n}\boldsymbol{A}\boldsymbol{x}$ where $a_n := \mathbb{E}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{y}\|_2] \approx \sqrt{n}$, see Lemma 1.1. Then for each fixed $\boldsymbol{x} \in S^{k-1}$ we have $\mathbb{E}_{\boldsymbol{A}}[\|T(\boldsymbol{x})\|_2] = 1$. While some deviation is possible we will be able to argue that $T(B_2^k)$ is approximately $S^{n-1} \cap F$.

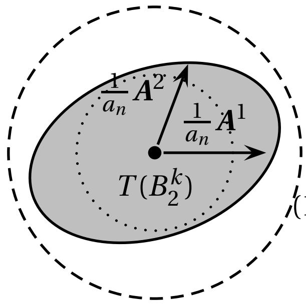
First, we prove a little helper lemma:

Lemma 9.30. *Let $T : \mathbb{R}^k \to \mathbb{R}^n$ be an linear map, let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $0 < \varepsilon \leq 1$ and $r > 0$. Consider the subspace $F := \{T(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^k\}$. If $|\|T(\boldsymbol{x})\|_K - r| \leq \varepsilon r \; \forall \boldsymbol{x} \in S^{k-1}$, then $(1-\varepsilon)r \cdot (K \cap F) \subseteq T(B_2^k) \subseteq (1+\varepsilon)r \cdot (K \cap F)$.*

Proof. After scaling assume $r = 1$. By convexity, for the 2nd inclusion it suffices to prove that $T(S^{k-1}) \subseteq (1+\varepsilon)K$. And indeed for $\boldsymbol{x} \in S^{k-1}$ one has $\|T(\boldsymbol{x})\|_K \leq (1+\varepsilon)$ by assumption. Now suppose that there is a $\boldsymbol{y} \in F$ with $\|\boldsymbol{y}\|_K < 1-\varepsilon$ and $\boldsymbol{y} \notin T(B_2^k)$. We can scale $\boldsymbol{y}$ down until $\boldsymbol{y} \in \partial T(B_2^k)$ and still $\|\boldsymbol{y}\|_K < 1-\varepsilon$. But if $\boldsymbol{y}$ lies on the boundary of $T(B_2^k)$ then also $\boldsymbol{y} \in T(S^{k-1})$. Again the assumption tells us that $|\|\boldsymbol{y}\|_K - 1| \leq \varepsilon$ which is a contradiction. □

Lemma 9.31. *There is a constant $c_0 > 0$ so that for $n \in \mathbb{N}$ and $\varepsilon > 0$ and $k \leq c_0\varepsilon^2 n$ the following holds: Draw $\boldsymbol{A} \sim N^{n\times k}(0,1)$ and consider the map $T : \mathbb{R}^k \to \mathbb{R}^n$ with $T(\boldsymbol{x}) := \frac{1}{a_n}\boldsymbol{A}\boldsymbol{x}$ and $F :=$ span$\{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^k\}$. Then*

$$\Pr\left[(1-\varepsilon)\cdot(B_2^n \cap F) \subseteq T(B_2^k) \subseteq (1+\varepsilon)(B_2^n \cap F)\right] \geq 0.99.$$



Proof. We apply the Matrix Concentration Theorem 9.29 and

$$\mathbb{E}\Big[\sup_{\boldsymbol{x}\in S^{k-1}}\Big|\|\boldsymbol{A}\boldsymbol{x}\|_2 - \underbrace{\mathbb{E}[\|\boldsymbol{A}\boldsymbol{x}\|_2]}_{=a_n}\Big|\Big] \leq O(1)\cdot \underbrace{g(S^{k-1})}_{=a_k} \leq O(\sqrt{k})$$

Then dividing by a_n and choosing the constant c_0 small enough gives

$$\mathbb{E}\Big[\sup_{\boldsymbol{x}\in S^{k-1}}\Big|\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{a_n} - 1\Big|\Big] \leq O(\sqrt{k/n}) \leq \frac{\varepsilon}{100}$$

Then with probability at least 0.99 one has

$$\sup_{\boldsymbol{x}\in S^{k-1}}\Big|\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{a_n} - 1\Big| \leq \varepsilon \quad (*)$$

Using Lemma 9.30 this then shows the claim. ☐

Now we come to the tight version of Dvoretzky's theorem which was proven first by Gordon.

Theorem 9.32 (Gordon [Gor85])**.** *Let $\varepsilon > 0$. For every symmetric convex body $K \subseteq \mathbb{R}^n$ there is a subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) = k =: \Theta(\varepsilon^2 \log(n))$ so that $d_{BM}(K\cap F, B_2^k) \leq 1+\varepsilon$.*

We already know that for any symmetric convex body K we can find a $\Omega(\log n)$-dimensional 2-spherical section. Then after scaling it only remains to prove the following:

Theorem 9.33. *For a universal constant $c_0 > 0$ the following holds: Let K be a symmetric convex body with $\frac{1}{2}B_2^n \subseteq K \subseteq 2B_2^n$ and let $0 < \varepsilon \leq 1$. Then there is a subspace $F \subseteq \mathbb{R}^n$ with $\dim(F) = k := c_0\varepsilon^2 n$ so that*

$$(1-\varepsilon)\cdot(B_2^n \cap F) \subseteq K \cap F \subseteq (1+\varepsilon)\cdot(B_2^n \cap F)$$

Proof. We draw a Gaussian random matrix $\boldsymbol{A} \sim N^{k\times n}(0,1)$ and consider the subspace $F := \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^k\}$ spanned by the columns of $\boldsymbol{A}$. Note that with probability 1 one has $\dim(F) = k$. We define a map $T : \mathbb{R}^k \to \mathbb{R}^n$ with $T(\boldsymbol{x}) := \frac{1}{a_n}\boldsymbol{A}\boldsymbol{x}$. Note that for a fixed $\boldsymbol{x} \in S^{k-1}$ one has $\mathbb{E}_{\boldsymbol{A}}[\|\boldsymbol{A}\boldsymbol{x}\|_K] = \mathbb{E}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[\|\boldsymbol{y}\|_K] = a_n \cdot \mathbb{E}_{\boldsymbol{y}\sim S^{n-1}}[\|\boldsymbol{y}\|_K] = a_n \cdot M(K)$. Then using Theorem 9.29 we get

$$\mathbb{E}\Big[\sup_{\boldsymbol{x}\in S^{k-1}} \Big|\|\boldsymbol{A}\boldsymbol{x}\|_K - \underbrace{\mathbb{E}\big[\|\boldsymbol{A}\boldsymbol{x}\|_K\big]}_{=a_n M(K)}\Big|\Big] \leq O(1)\cdot \underbrace{g(S^{k-1})}_{=a_k}$$

Then by Markov's inequality the left hand side expectation is at most 100 times the right hand side with probability 99%. We fix a matrix $\boldsymbol{A}$ where this indeed happens and the event from Lemma 9.31 happens. Then dividing by a_n gives

$$\sup_{\boldsymbol{x}\in S^{k-1}} \Big|\|T(\boldsymbol{x})\|_K - M(K)\Big| \leq O\Big(\frac{a_k}{a_n}\Big) = \Theta\Big(\sqrt{k/n}\Big) \leq \frac{\varepsilon}{3}M(K) \quad (*)$$

using $M(K) \geq 1/2$ as $K \subseteq 2B_2^n$ and using that $k = c_0\varepsilon^2 n$ with a small enough constant c_0. Then combining $(*)$ with Lemma 9.30 we know

$$\Big(1-\frac{\varepsilon}{3}\Big)\cdot M(K)\cdot(K\cap F) \subseteq T(B_2^k) \subseteq \Big(1+\frac{\varepsilon}{3}\Big)\cdot M(K)\cdot(K\cap F) \qquad (**)$$

and from Lemma 9.31 we know that also

$$\Big(1-\frac{\varepsilon}{3}\Big)\cdot(B_2^n\cap F) \subseteq T(B_2^k) \subseteq \Big(1+\frac{\varepsilon}{3}\Big)\cdot(B_2^n\cap F) \qquad (***)$$

Combining $(**)$ and $(***)$ we obtain

$$\underbrace{\frac{1-\frac{\varepsilon}{3}}{1+\frac{\varepsilon}{3}}}_{\geq 1-\varepsilon}\cdot(B_2^n\cap F) \subseteq M(K)\cdot(K\cap F) \subseteq \underbrace{\frac{1+\frac{\varepsilon}{3}}{1-\frac{\varepsilon}{3}}}_{\leq 1+\varepsilon}\cdot(B_2^n\cap F)$$

as desired. □

9.9 Exercises

Exercise 9.1.
Let $T \subseteq \mathbb{R}^n$. Prove that if T is bounded, then for any $\varepsilon > 0$ there is a finite set $T' \subseteq T$ with $g(T) - \varepsilon \leq g(T') \leq g(T)$.

Exercise 9.2.
Consider the Gaussian process $\mathcal{X} = (X_t)_{t\in\mathbb{N}}$ of independent Gaussians with $\mathbb{E}[X_t^2]^{1/2} = \frac{1}{\sqrt{1+\log_2(t)}}$).

(a) Prove that $\mathbb{E}[\sup_{t\in\mathbb{N}} X_t]$ is finite.

(b) Prove that Dudley's upper bound is unbounded[6].

(c) Give an admissible sequence which shows that $\gamma_2(\mathcal{X})$ is finite.

Exercise 9.3.
Prove that for any symmetric set $T \subseteq \mathbb{R}^n$ and any symmetric convex body $K \subseteq \mathbb{R}^m$ one has

$$\mathop{\mathbb{E}}_{\boldsymbol{A}\sim N^{m\times n}(0,1)}\left[\sup_{\boldsymbol{x}\in T}\left|\sup_{\boldsymbol{y}\in K}\langle \boldsymbol{A}\boldsymbol{x},\boldsymbol{y}\rangle - g(K)\cdot\|\boldsymbol{x}\|_2\right|\right] \le O(g(T))\cdot \mathrm{radius}(K)$$

Hint. Use Theorem 9.29.
Remark. This is known as "*Two-sided Chevet's Inequality*" or "*General Chevet Inequality*".

Exercise 9.4.
Let $T \subseteq S^{n-1}$ be a symmetric set and draw $\boldsymbol{A} \sim N^{m\times n}$ where $g(T) \le C_0\sqrt{m}$ for a small enough constant C_0. Then with probability at least $1/2$ one has $T \cap \ker(\boldsymbol{A}) = \emptyset$.
Hint: Apply Theorem 9.29.
Remark: This is known as the "*Escape through a Mesh Theorem*".

Exercise 9.5.
Let $S \subseteq S^{n-1}$ be a finite set of points. Let $k := \frac{C_0}{\varepsilon^2}\log(|S|)$ for a large enough constant C_0. Show that for a Gaussian random matrix $\boldsymbol{A} \sim N^{k\times n}(0,1)$ with probability at least $1/2$ one has $(1-\varepsilon)a_k\|\boldsymbol{x}-\boldsymbol{y}\|_2 \le \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}\boldsymbol{y}\|_2 \le (1+\varepsilon)a_k\|\boldsymbol{x}-\boldsymbol{y}\|_2$ for all $\boldsymbol{x},\boldsymbol{y} \in S$.
Hint. Apply Theorem 9.29.
Remark. This is known as "*Johnson-Lindenstrauss Lemma*".

Exercise 9.6.
Prove that for a large enough constant C the following holds: A random matrix $\boldsymbol{A} \sim N^{m\times n}(0,1)$ with $m \ge Cn$ has $\frac{1}{2}\sqrt{m} \le \sigma_i(\boldsymbol{A}) \le 2\sqrt{m}$ for all $i \in [n]$ with probability at least $1/2$.
Hint. Recall that $\sigma_i(\boldsymbol{A})$ gives the ith largest singular value of $\boldsymbol{A}$. In particular $\sigma_1(\boldsymbol{A}) = \max_{\boldsymbol{x}\in S^{n-1}}\|\boldsymbol{A}\boldsymbol{x}\|_2$ and $\sigma_n(\boldsymbol{A}) = \min_{\boldsymbol{x}\in S^{n-1}}\|\boldsymbol{A}\boldsymbol{x}\|_2$. Use again Theorem 9.29.

[6] In the proof of Dudley's upper bound we have made the assumption of a finite Gaussian process but the bound indeed holds in the infinite case too.

Chapter 10

Volume distribution in convex bodies and the isotropic position

In this chapter we will discuss the distribution of volume in a convex body. For this sake we are introducing another "standard position" for a convex body where we ask that the mass of a body is as normalized as possible. This position will turn out to be helpful for many types of such volume considerations.

10.1 Introduction

Before we begin, recall that we say a body K is *centered*, if $\text{bary}(K) = \mathbb{E}_{\boldsymbol{x}\sim K}[\boldsymbol{x}] = \boldsymbol{0}$. The central definition is as follows:

Definition 10.1. A convex body $K \subseteq \mathbb{R}^n$ is in *isotropic position*, if the following 3 conditions are satisfied:

(I) $\text{Vol}_n(K) = 1$.
(II) The barycenter of K is the origin.
(III) One has $\mathbb{E}_{\boldsymbol{x}\sim K}\left[\boldsymbol{x}\boldsymbol{x}^T\right] = \alpha \cdot \boldsymbol{I}_n$ for some $\alpha \geq 0$.

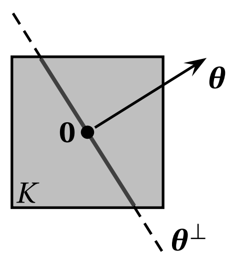
Another way to see this is that if we sample $\boldsymbol{x} \sim K$ uniformly, then the 1st and 2nd moment of $\boldsymbol{x}$ are identical to a scaled standard Gaussian. The constant α from (III) is going to be of particular interest to us:

Definition 10.2. For a convex body $K \subseteq \mathbb{R}^n$ that is in isotropic position we define the *isotropic constant* as the number $L_K \in \mathbb{R}_{\geq 0}$ satisfying $\mathbb{E}_{\boldsymbol{x}\sim K}\left[\boldsymbol{x}\boldsymbol{x}^T\right] = L_K^2 \cdot \boldsymbol{I}_n$. For a convex body K that may not be in isotropic position we set $L_K := L_{T(K)}$ where T is a volume-preserving affine map[1] so that $T(K)$ is in isotropic position.

[1]Recall that $T : \mathbb{R}^n \to \mathbb{R}^n$ with $T(\boldsymbol{x}) := \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with $\boldsymbol{A} \in \mathbb{R}^{n\times n}, \boldsymbol{b} \in \mathbb{R}^n$ and $|\det(\boldsymbol{A})| = 1$ is affine and volume-preserving in the sense that that $\text{Vol}_n(T(K)) = \text{Vol}_n(K)$ for any measurable set K.

We will see later in Sec 10.3 that the map T putting $T(K)$ in isotropic position is unique up to rotations and so this definition is well-defined. It is worth noting that for an isotropic convex body K one has $L_K = \frac{\mathbb{E}_{\boldsymbol{x}\sim K}[\|\boldsymbol{x}\|_2^2]^{1/2}}{\sqrt{n}}$ meaning that L_K is somewhat proportional to the average length of points in K. We will see that always $L_K \geq \Omega(1)$ and L_K is indeed minimized for the Euclidean ball (see Sec 10.4). On the other hand, upper bounds on L_K are a lot less understood. It is easy to derive from John's Theorem that at least for symmetric convex bodies one has $L_K \leq O(\sqrt{n})$ and with a bit more work one can prove that $L_K \leq O(n^{1/4}\log(n))$, which is due to Bourgain (see Section 10.8). But it is indeed conjectured that a constant upper bound on L_K is possible. This can be phrased as a conjecture in various different forms:

- **Isotropic Constant Conjecture I.** One has $L_K \leq O(1)$ for any convex body K.
- **Isotropic Constant Conjecture II.** For any convex body $K \subseteq \mathbb{R}^n$ in isotropic position and any direction $\boldsymbol{y} \in S^{n-1}$ one has $\mathbb{E}_{\boldsymbol{x}\sim K}[\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2] \leq O(1)$.
- **Slicing Conjecture I.** For any convex body K in isotropic position and any direction $\boldsymbol{\theta} \in S^{n-1}$ one has $\mathrm{Vol}_{n-1}(K \cap \boldsymbol{\theta}^\perp) \geq \Omega(1)$.[2]
- **Slicing Conjecture II.** For any centered convex body K with $\mathrm{Vol}_n(K) = 1$, there exists at least <u>one</u> direction $\boldsymbol{\theta} \in S^{n-1}$ so that $\mathrm{Vol}_{n-1}(K \cap \boldsymbol{\theta}^\perp) \geq \Omega(1)$.



We will later give formal statements and prove that in fact all these conjectures are equivalent. The chapter will then close with a proof of Bourgain's upper bound of $L_K \leq O(n^{1/4}\log(n))$; a key estimate that we need towards that goal is that for any convex body K in isotropic position and any direction $\boldsymbol{\theta}$ one has $\mathrm{Vol}_n(\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle \geq t \cdot L_K\}) \leq \exp(-\Theta(t))$ for any $t \geq \Theta(1)$. In particular from this one can see that the isotropic constant L_K controls how much volume non-centered slices of K may contain.

[2]We abbreviate $\boldsymbol{\theta}^\perp := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \perp \boldsymbol{\theta}\}$ as the $(n-1)$-dimensional subspace that is orthogonal to the vector $\boldsymbol{\theta}$.

10.2 The basics

To get some intuition we want to discuss at least one concrete example. Consider the *cube* $K := [-\frac{1}{2}, \frac{1}{2}]^n$ which has $\mathrm{Vol}_n(K) = 1$ and is centered. Moreover $\mathbb{E}_{\boldsymbol{x}\sim K}[\boldsymbol{x}\boldsymbol{x}^T] = \alpha \cdot \boldsymbol{I}_n$ where $\alpha = \mathbb{E}_{x\sim[-\frac{1}{2},\frac{1}{2}]}[x^2] = \frac{1}{12}$. Hence we know that K is in isotropic position and $L_K = \frac{1}{\sqrt{12}}$. Next we will formalize a claim that we made earlier:

Lemma 10.3. *If $K \subseteq \mathbb{R}^n$ is in isotropic position then $\mathbb{E}_{\boldsymbol{x}\sim K}[\|\boldsymbol{x}\|_2^2] = nL_K^2$ and $\mathbb{E}_{\boldsymbol{x}\sim K}[\|\boldsymbol{x}\|_2] \leq \sqrt{n} \cdot L_K$.*

Proof. Simply write

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\left[\|\boldsymbol{x}\|_2\right]^2 \overset{\text{Jensen}}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\left[\|\boldsymbol{x}\|_2^2\right] = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\mathrm{Tr}\big[\boldsymbol{x}\boldsymbol{x}^T\big]\Big] \overset{\text{linearity}}{=} \mathrm{Tr}\Big[\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\big[\boldsymbol{x}\boldsymbol{x}^T\big]}_{=L_K^2\boldsymbol{I}_n}\Big] = L_K^2 n$$

where Jensen's inequality applies as $y \mapsto y^2$ is a convex function. □

We show a useful lemma which implies that in order to check isotropy of a body it suffices to verify a 1-dimensional integral.

Lemma 10.4. *Suppose that $K \subseteq \mathbb{R}^n$ is convex with bary$(K) = \boldsymbol{0}$ and Vol$_n(K) = 1$. Then the following is equivalent:*

- *K is isotropic with constant L_K.*
- *One has $\mathbb{E}_{\boldsymbol{x}\sim K}[\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2] = L_K^2\|\boldsymbol{y}\|_2^2$ for every $\boldsymbol{y} \in \mathbb{R}^n$.*

Proof. Abbreviate $\boldsymbol{M} := \mathbb{E}_{\boldsymbol{x}\sim K}[\boldsymbol{x}\boldsymbol{x}^T]$. Then $\mathbb{E}_{\boldsymbol{x}\sim K}[\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2] = \langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle$. We can now see that $\langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle = L_K^2\|\boldsymbol{y}\|_2^2\ \forall \boldsymbol{y} \in \mathbb{R}^n$ iff $\boldsymbol{M} = L_K^2\boldsymbol{I}_n$. □

Phrased differently, for any convex body K in isotropic position, the random variable $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ with $\boldsymbol{x} \sim K$ has a *standard deviation* of exactly L_K for any direction $\boldsymbol{y} \in S^{n-1}$. We want to restate the first two conjectures that we mentioned in the introduction:

Conjecture 1 (Isotropic Constant Conjecture I)**.** *There is a universal constant $C > 0$ so that $L_K \leq C$ for any convex body $K \subseteq \mathbb{R}^n$.*

Conjecture 2 (Isotropic Constant Conjecture II)**.** *There is an absolute constant $C > 0$ so that every convex body $K \subseteq \mathbb{R}^n$ in isotropic position satisfies*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}[\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2] \leq C^2 \quad \forall \boldsymbol{y} \in S^{n-1}$$

Then from Lemma 10.4 we see that these two conjectures are equivalent and indeed the best achievable values of C are the same in both.

10.3 Existence and uniqueness of isotropic position

It still remains to justify that any convex body K can be brought into isotropic position and that the position is unique up to rotation.

Lemma 10.5. *Let $K \subseteq \mathbb{R}^n$ be a centered convex body. Then there exists a bijective linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $T(K)$ is isotropic.*

Proof. Consider the matrix $\boldsymbol{M} := \int_K \boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x}$. Then by construction, $\boldsymbol{M}$ is symmetric and positive definite and hence it has a square root $\boldsymbol{M}^{1/2} \succ 0$. Consider the map $T(\boldsymbol{x}) := \boldsymbol{M}^{-1/2}\boldsymbol{x}$. Then we can verify that for every $\boldsymbol{y} \in \mathbb{R}^n$ one has

$$\begin{aligned}
\int_{T(K)} \langle \boldsymbol{x}, \boldsymbol{y}\rangle^2 d\boldsymbol{x} &= \det(\boldsymbol{M}^{-1/2}) \int_K \langle \boldsymbol{M}^{-1/2}\boldsymbol{x}, \boldsymbol{y}\rangle^2 d\boldsymbol{x} \\
&= \det(\boldsymbol{M}^{-1/2}) \int_K \langle \boldsymbol{x}\boldsymbol{x}^T, (\boldsymbol{M}^{-1/2}\boldsymbol{y})(\boldsymbol{M}^{-1/2}\boldsymbol{y})^T\rangle \, d\boldsymbol{x} \\
&= \det(\boldsymbol{M}^{-1/2}) \cdot \mathrm{Tr}\big[\boldsymbol{M}\boldsymbol{M}^{-1/2}\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{M}^{-1/2}\big] = \det(\boldsymbol{M}^{-1/2}) \cdot \|\boldsymbol{y}\|_2^2
\end{aligned}$$

using an integral transformation. Then a scaling of $T(K)$ will be in isotropic position. □

Next, we can prove that the isotropic position is *unique* up to orthogonal transformations. In the same proof we will learn that the isotropic position arises as a solution to a minimization problem. To be more precise, the isotropic position will be the one that minimizes the *average* $\|\cdot\|_2^2$*-length* of points in K.

Lemma 10.6. *Let $K \subseteq \mathbb{R}^n$ be a centered convex body with $\mathrm{Vol}_n(K) = 1$. Define*

$$B(K) := \inf\Big\{ \int_{T(K)} \|\boldsymbol{x}\|_2^2 \, d\boldsymbol{x} \mid T : \mathbb{R}^n \to \mathbb{R}^n \text{ linear map with } |\det(\boldsymbol{T})| = 1 \Big\}$$

Then a position $K_1 = T(K)$ (with $\mathrm{Vol}_n(K_1) = 1$) is isotropic if and only if

$$\int_{T(K)} \|\boldsymbol{x}\|_2^2 \, d\boldsymbol{x} = B(K)$$

Moreover, if K_1 and K_2 are isotropic positions of K, then $K_2 = U(K_1)$ for some orthogonal transformation $U : \mathbb{R}^n \to \mathbb{R}^n$.

Proof. Let K_1 be an isotropic position of K. Then $\int_{K_1} \boldsymbol{x}\boldsymbol{x}^T \, d\boldsymbol{x} = \alpha \cdot \boldsymbol{I}_n$ for some $\alpha > 0$ and $\int_{K_1} \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} = \mathrm{Tr}[\int_{K_1} \boldsymbol{x}\boldsymbol{x}^T \, d\boldsymbol{x}] = \alpha n$. We need to prove that no other transformation could achieve a smaller value.

For this purpose, fix any linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ with $|\det(\boldsymbol{A})| = 1$ and estimate that

$$\int_{A(K_1)} \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} = \int_{K_1} \|\boldsymbol{A}\boldsymbol{x}\|_2^2 d\boldsymbol{x} = \int_{K_1} \mathrm{Tr}\big[\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\big] d\boldsymbol{x} \overset{\text{linearity}}{=} \alpha \cdot \mathrm{Tr}[\boldsymbol{A}^T\boldsymbol{A}] \geq \alpha n$$

here we use that $\boldsymbol{A}^T\boldsymbol{A}$ is a symmetric PSD matrix with $\det(\boldsymbol{A}^T\boldsymbol{A}) = 1$ and so $\mathrm{Tr}[\boldsymbol{A}^T\boldsymbol{A}] \geq n$ by the *arithmetic-geometric mean inequality*. Moreover, the only case where one has equality is if $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}_n$ which means that $\boldsymbol{A}$ is an orthogonal matrix. That concludes the uniqueness proof. ☐

We can also give an alternative variational argument. We restate the part of the claim that we reprove:

Lemma 10.7. *Suppose that K is a centered convex body with $\mathrm{Vol}_n(K) = 1$ and the property that $\boldsymbol{I}_n$ is an optimum for the minimization problem. Then K is in isotropic position.*

Proof. Take any linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ with $\det(\boldsymbol{A}) = 1$. Let $\varepsilon > 0$ be small enough so that $\boldsymbol{I}_n + \varepsilon\boldsymbol{A}$ is invertible and consider the matrix $\boldsymbol{B} := \frac{\boldsymbol{I}_n+\varepsilon\boldsymbol{A}}{\det(\boldsymbol{I}_n+\varepsilon\boldsymbol{A})^{1/n}}$. Then by minimality

$$\begin{aligned}
\int_K \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} &\leq \int_K \|\boldsymbol{B}\boldsymbol{x}\|_2^2 d\boldsymbol{x} = \frac{1}{\det(\boldsymbol{I}_n + \varepsilon\boldsymbol{A})^{2/n}} \int_K \|\boldsymbol{x} + \varepsilon\boldsymbol{A}\boldsymbol{x}\|_2^2 \, d\boldsymbol{x} \\
&= \frac{1}{1 + \frac{2\varepsilon}{n}\mathrm{Tr}[\boldsymbol{A}] + O(\varepsilon^2)} \int_K \Big(\|\boldsymbol{x}\|_2^2 + 2\varepsilon \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle + O(\varepsilon^2)\Big) d\boldsymbol{x}
\end{aligned}$$

Then rearranging, and comparing the derivative w.r.t. ε at $\varepsilon = 0$ (which makes the $O(\varepsilon^2)$ terms disappear) we obtain

$$\frac{1}{n}\mathrm{Tr}[\boldsymbol{A}] \cdot \int_K \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} \leq \int_K \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle \, d\boldsymbol{x} \quad (*)$$

Repeating the argument with $-\boldsymbol{A}$ we obtain that the inequality $(*)$ holds with equality; after rearranging we obtain

$$\int_K \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} = \frac{n}{\mathrm{Tr}[\boldsymbol{A}]} \int_K \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle \, d\boldsymbol{x} = \frac{n}{\mathrm{Tr}[\boldsymbol{A}]} \cdot \langle \int_K \boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x}, \boldsymbol{A}\rangle$$

for all invertible linear maps $\boldsymbol{A}$. The only way that the right hand side expression is constant over all $\boldsymbol{A}$ is if $\int_K \boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x} = \alpha\boldsymbol{I}_n$ for some α, which means K was in isotropic position. ☐

The discussion above leads to an alternative characterization of L_K (which also could be used as a definition):

Lemma 10.8. *Let $K \subseteq \mathbb{R}^n$ be a convex body with bary$(K) = \mathbf{0}$. Then*

$$L_K = \Big(\frac{1}{n}\min\Big\{\frac{\mathbb{E}_{\boldsymbol{x}\sim T(K)}[\|\boldsymbol{x}\|_2^2]}{Vol_n(T(K))^{2/n}} \mid T \textit{ bijective linear map}\Big\}\Big)^{1/2}$$

Proof. Applying a linear transformation to K does obviously not change the right hand side, hence we may as well assume that K is in isotropic position. Then by Lemma 10.6 the minimum is attained for the identity $T(\boldsymbol{x}) = \boldsymbol{x}$. Then

$$\Big(\frac{1}{n}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}[\|\boldsymbol{x}\|_2^2]\Big)^{1/2} \stackrel{\text{Cor 10.3}}{=} \Big(\frac{1}{n}\cdot nL_K^2\Big)^{1/2} = L_K$$

as claimed. $\square$

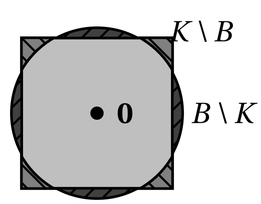
This characterization is useful because for any centered convex body $K \subseteq \mathbb{R}^n$ and any linear map T with $\mathrm{Vol}_n(T(K)) = 1$ we obtain an *upper bound* of $L_K \leq \frac{\mathbb{E}_{\boldsymbol{x}\sim T(K)}[\|\boldsymbol{x}\|_2^2]^{1/2}}{\sqrt{n}}$.

10.4 Lower and upper bounds on L_K

Interestingly one can prove that the Euclidean ball is the body that minimizes the isotropic constant and hence $L_K \geq \Omega(1)$ for any convex body K.

Lemma 10.9. *Every convex body $K \subseteq \mathbb{R}^n$ has $L_K \geq L_{B_2^n} \geq c$, where $c > 0$ is an absolute constant.*

Proof. The isotropic constant L_K is invariant under linear transformations, so me may assume that K is in isotropic position. We pick r so that the ball $B := rB_2^n$ has $\mathrm{Vol}_n(B) = 1$. Note that also B is isotropic. Observe that $\mathrm{Vol}_n(K \setminus B) = \mathrm{Vol}_n(B \setminus K)$, but the vectors in $K \setminus B$ are longer than the ones in $B \setminus K$.



This insight alone implies the statement as "moving" mass from $K \setminus B$ to $B \setminus K$ can only decrease the average $\|\cdot\|_2^2$-length. More formally one can verify that

$$\begin{aligned} L_K &\overset{K \text{ isotropic}}{=} \Big(\frac{1}{n}\int_K \|\boldsymbol{x}\|_2^2 d\boldsymbol{x}\Big)^{1/2} = \Big(\frac{1}{n}\int_{K\cap B} \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} + \frac{1}{n}\int_{K\setminus B} \|\boldsymbol{x}\|_2^2 d\boldsymbol{x}\Big)^{1/2} \\ &\geq \Big(\frac{1}{n}\int_{K\cap B} \|\boldsymbol{x}\|_2^2 d\boldsymbol{x} + \frac{1}{n}\int_{B\setminus K} \|\boldsymbol{x}\|_2^2 d\boldsymbol{x}\Big)^{1/2} \overset{B \text{ isotropic}}{=} L_B \end{aligned}$$

We have already justfied earlier that $L_{B_2^n} \geq c$ for some universal constant $c > 0$. $\square$

In fact, there is no convex body K known where even $L_K > 1$. We can obtain a first rather weak upper bound on L_K using John's Theorem. Recall that $d_{BM}(K)$ is the Banach-Mazur distance of a body to B_2^n.

Lemma 10.10. *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric convex body. Then $L_K \leq O(d_{BM}(K)) \leq O(\sqrt{n})$.*

Proof. Apply a linear transformation to K so that $\mathrm{Vol}_n(K) = 1$ and $rB_2^n \subseteq K \subseteq r\,d_{BM}(K)B_2^n$. Clearly $r \leq \Theta(\sqrt{n})$ because $\mathrm{Vol}_n(\Theta(\sqrt{n}) \cdot B_2^n) = 1$. We cannot guarantee that K is now in isotropic position, but still by Lemma 10.8 we can obtain an upper bound of

$$L_K \leq \Big(\frac{1}{n}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}[\|\boldsymbol{x}\|_2^2]\Big)^{1/2} \leq \Big(\frac{1}{n}(r \cdot d_{BM}(K))^2\Big)^{1/2} \leq O(d_{BM}(K))$$

$\square$

One can prove that in order to solve the isotropic constant conjecture one can restrict the attention to symmetric bodies. The proof is a bit more involved and we simply state the result which uses a construction of Keith Ball:

Theorem 10.11. *For every convex body $K \subseteq \mathbb{R}^n$, there exists a centrally symmetric convex body $Q \subseteq \mathbb{R}^n$ so that $L_K \leq C \cdot L_Q$ where $C > 0$ is a universal constant.*

10.5 Moments of Inertia and maximal hyperplane sections

Let $K \subseteq \mathbb{R}^n$ be a centered convex body with $\mathrm{Vol}_n(K) = 1$. We define the *matrix of inertia of K* as

$$\boldsymbol{M}_K := \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\big[\boldsymbol{x}\boldsymbol{x}^T\big]$$

Then $\boldsymbol{M}_K$ is a positive definite matrix by construction and hence there is a positive definite *square root* $\boldsymbol{M}_K^{1/2}$. As usually we also denote M_K and $M_K^{-1/2}$ as the linear maps that are described by the matrices $\boldsymbol{M}_K$ and $\boldsymbol{M}_K^{-1/2}$. We define the *Binet ellipsoid* of K as

$$\mathcal{E}_B(K) := M_K^{-1/2}(B_2^n) = \{\boldsymbol{M}_K^{-1/2}\boldsymbol{x} \mid \boldsymbol{x} \in B_2^n\}$$

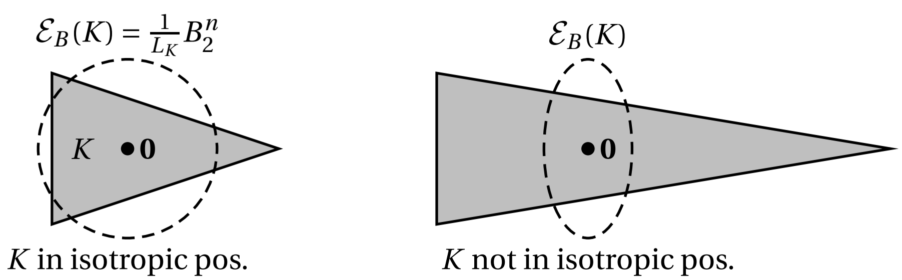
Then the norm induced by that ellipsoid satisfies

$$\|\boldsymbol{y}\|_{\mathcal{E}_B(K)} = \|\boldsymbol{M}_K^{1/2}\boldsymbol{y}\|_2 = (\boldsymbol{y}^T\boldsymbol{M}_K\boldsymbol{y})^{1/2} = \Big(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\big[\langle \boldsymbol{x},\boldsymbol{y}\rangle^2\big]\Big)^{1/2}$$

This means that for any $\boldsymbol{y} \in S^{n-1}$, the standard deviation of the random variable $\langle \boldsymbol{x},\boldsymbol{y}\rangle$ with $\boldsymbol{y} \sim K$ is given by $\|\boldsymbol{y}\|_{\mathcal{E}_B(K)}$. From what we know already it follows that:

Corollary 10.12. *Let $K \subseteq \mathbb{R}^n$ be a centered convex body. Then K is in isotropic position if and only if $\mathcal{E}_B(K) = L_K^{-1} \cdot B_2^n$.*

Geometrically one may imagine that $\mathcal{E}_B(K)$ behaves "inverse proportionally" to K:



As we mentioned earlier, a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ with $|\det(\boldsymbol{T})| = 1$ is volume preserving. We will now prove that $\mathrm{Vol}_n(\mathcal{E}_B(K)) = \mathrm{Vol}_n(\mathcal{E}_B(T(K))$ by showing that $\mathcal{E}_B(K)$ is a function of the isotropic constant L_K.

Lemma 10.13. *Let $K \subseteq \mathbb{R}^n$ be a centered convex body with $Vol_n(K) = 1$. Then*

$$Vol_n(\mathcal{E}_B(K)) = Vol_n(B_2^n) \cdot L_K^{-n}$$

Proof. One can easily check that for any bijective linear map $A : \mathbb{R}^n \to \mathbb{R}^n$, the matrix of inertia changes as $\boldsymbol{M}_{A(K)} = \boldsymbol{A}\boldsymbol{M}_K\boldsymbol{A}^T$ when applying a linear transformation. Then let A be the map with $|\det(\boldsymbol{A})| = 1$ so that $A(K)$ is in isotropic position and so $\mathcal{E}_B(A(K)) = L_K^{-1}B_2^n$. We know that $\boldsymbol{M}_{A(K)} = \boldsymbol{A}\boldsymbol{M}_K\boldsymbol{A}^T$, but these two matrices have identical determinant. Then

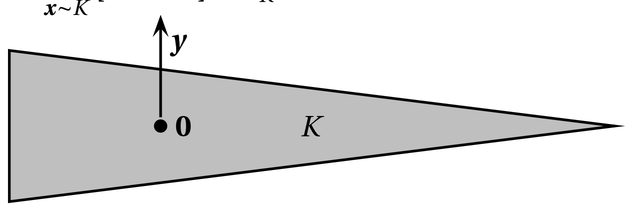
$$\begin{aligned} L_K^{-n}\mathrm{Vol}_n(B_2^n) &\overset{A(K)\text{ in isot. pos}}{=} \mathrm{Vol}_n(\mathcal{E}_B(A(K))) \\ &\overset{\text{Vol. of ellipsoid}}{=} \mathrm{Vol}_n(B_2^n)\cdot|\det(\boldsymbol{M}_{A(K)})|^{-1/2} \\ &= \mathrm{Vol}_n(B_2^n)\cdot|\det(\boldsymbol{M}_K)|^{-1/2} \\ &\overset{\text{Vol. of ellipsoid}}{=} \mathrm{Vol}_n(\mathcal{E}_B(K)) \end{aligned}$$

□

We can now derive that every convex body has a "thin direction" depending on its isotropic constant. Interestingly for this statement, the body K does not need to be in isotropic position.

Lemma 10.14. *Let $K \subseteq \mathbb{R}^n$ be a centered convex body with $\mathrm{Vol}_n(K) = 1$. Then there exists a direction $\boldsymbol{y} \in S^{n-1}$ so that*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\left[\langle \boldsymbol{x},\boldsymbol{y}\rangle^2\right] \leq L_K^2$$



Proof. Recall that for a symmetric convex body Q, the radius in direction $\boldsymbol{y}$ is denoted by $\rho_Q(\boldsymbol{y}) = \max\{r \geq 0 \mid r\boldsymbol{y} \in Q\}$. Integrating in polar coordinates (see Lem 1.46) gives

$$\mathop{\mathbb{E}}_{\boldsymbol{y}\sim S^{n-1}}\left[\|\boldsymbol{y}\|_Q^{-n}\right] = \mathop{\mathbb{E}}_{\boldsymbol{y}\sim S^{n-1}}\left[\rho_Q(\boldsymbol{y})^n\right] = \frac{\mathrm{Vol}_n(Q)}{\mathrm{Vol}_n(B_2^n)}$$

Applying this to the symmetric convex body $\mathcal{E}_B(K)$ gives

$$\mathop{\mathbb{E}}_{\boldsymbol{y}\sim S^{n-1}}\left[\|\boldsymbol{y}\|_{\mathcal{E}_B(K)}^{-n}\right] \overset{\substack{\text{integrating in}\\ \text{polar coord}}}{=} \frac{\mathrm{Vol}_n(\mathcal{E}_B(K))}{\mathrm{Vol}_n(B_2^n)} \overset{\text{Lem 10.13}}{=} L_K^{-n}$$

Hence there must be at least one vector $\boldsymbol{y}^* \sim S^{n-1}$ where indeed $\|\boldsymbol{y}^*\|_{\mathcal{E}_B(K)} \leq L_K$. We discussed earlier that $\|\boldsymbol{y}^*\|_{\mathcal{E}_B(K)} = \mathbb{E}_{\boldsymbol{x}\sim K}[\langle\boldsymbol{x},\boldsymbol{y}\rangle^2]^{1/2}$ which proves the claim. □

In the proof we used an average argument over an *inverse* quantity. And indeed for a long and skinny body such as $K := [-\frac{1}{2N}, \frac{1}{2N}] \times [\frac{N}{2}, \frac{N}{2}]$ almost all of the vectors $\boldsymbol{y}$ will have $\mathbb{E}_{\boldsymbol{x}\sim K}[\langle\boldsymbol{x},\boldsymbol{y}\rangle^2] \gg L_K^2$.

We want to eventually prove that for a convex body in isotropic position, every slice $K \cap \boldsymbol{\theta}^\perp$ has the same volume — up to a constant factor. For this goal, we need a couple of technical estimates.

10.6 The Maximum of Log concave functions

Next we show an important inequality that relates the maximum of a log concave function with its bary center. Recall that a function $f : \mathbb{R}^n \to [0,\infty[$ is *centered* if $\int_{\mathbb{R}^n} \boldsymbol{x} \cdot f(\boldsymbol{x}) d\boldsymbol{x} = \boldsymbol{0}$.

Theorem 10.15 (Fradelizi)**.** *Let* $f : \mathbb{R}^n \to [0,\infty[$ *be a centered log-concave function. Then* $\sup\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n\} \leq e^n \cdot f(\boldsymbol{0})$.

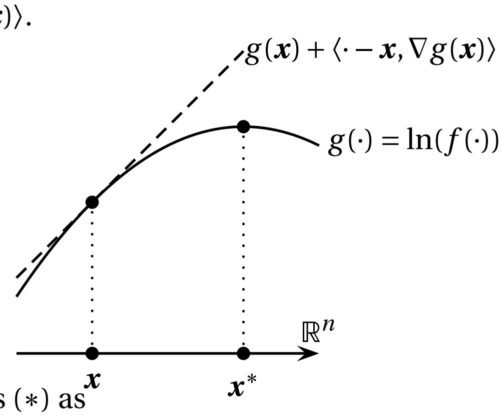
Proof. Let us scale the function so that $\int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x} = 1$. Then we can interpret f as the density function of some distribution μ. Note that the property that f is *centered* means that $\mathbb{E}_{\boldsymbol{y}\sim\mu}[\boldsymbol{x}] = \boldsymbol{0}$. We will show that for a fixed $\boldsymbol{x}^* \in \mathbb{R}^n$ one has

$$\ln(f(\boldsymbol{x}^*)) \overset{(*)}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\big[\ln(f(\boldsymbol{x}))\big] + n \overset{(**)}{\leq} \ln(f(\boldsymbol{0})) + n$$

In fact, we can see $(**)$ from

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\big[\ln(f(\boldsymbol{x}))\big] \overset{\text{Jensen}}{\leq} \ln\Big(f\Big(\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}]}_{=\boldsymbol{0}}\Big)\Big) = \ln(f(\boldsymbol{0}))$$

where we apply Jensen inequality with the fact that $\ln(f(\boldsymbol{x}))$ is concave. Recall that for a concave function $g(\boldsymbol{x}) := \ln(f(\boldsymbol{x}))$ we always have the upper bound $g(\boldsymbol{x}^*) \leq g(\boldsymbol{x}) + \langle \boldsymbol{x}^* - \boldsymbol{x}, \nabla g(\boldsymbol{x})\rangle$.



Then averaging over $\boldsymbol{x}$ gives $(*)$ as

$$\ln(f(\boldsymbol{x}^*)) \quad \leq \quad \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\Big[\ln(f(\boldsymbol{x}))\Big] + \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\Big[\langle \boldsymbol{x}^* - \boldsymbol{x}, \nabla g(\boldsymbol{x})\rangle\Big]}_{(***)}$$

It remains to show that $(***) = n$. Recall that $\nabla g(\boldsymbol{x}) = \frac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}$. Then switching from the average to the integral view gives

$$(***) = \int_{\mathbb{R}^n} f(\boldsymbol{x})\,\langle \boldsymbol{x}^* - \boldsymbol{x}, \frac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}\rangle\, d\boldsymbol{x} = \underbrace{\int_{\mathbb{R}^n} \langle \boldsymbol{x}^*, \nabla f(\boldsymbol{x})\rangle\, d\boldsymbol{x}}_{=0} - \underbrace{\int_{\mathbb{R}^n} \langle \boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle\, d\boldsymbol{x}}_{=-n} = n$$

Here we use the following:

Claim. *Let $f : \mathbb{R}^n \to [0,\infty[$ be an integrable function with* $\lim_{\|\boldsymbol{x}\|_2\to\infty} \|\boldsymbol{x}\|_2^2 f(\boldsymbol{x}) d\boldsymbol{x} = 0$ *and* $\int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x} = 1$*. Then* $\int_{\mathbb{R}^n} \langle \boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle \, d\boldsymbol{x} = -n$*.*

Proof of Claim. First consider a 1-dimensional function $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $\lim_{|x|\to\infty} |x \cdot h(x)| = 0$ (i.e. quickly decaying). *Integration by parts* gives $\int x \cdot h'(x) dx = x \cdot h(x) - \int h(x) dx$. So

$$\int_{\mathbb{R}} x \cdot h'(x) dx = \underbrace{[x \cdot h(x)]_{-\infty}^{\infty}}_{=0} - \int_{\mathbb{R}} h(x) dx.$$

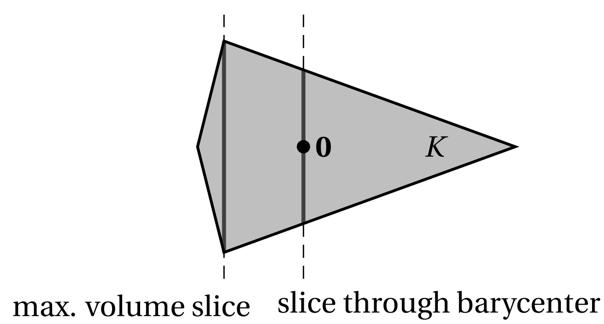
Now consider the function f. Then

$$\begin{aligned}
\int_{\mathbb{R}^n} \langle \boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle \, d\boldsymbol{x} &= \int_{\mathbb{R}^n} \Big(\sum_{i=1}^n x_i \frac{\partial}{\partial x_i} f(\boldsymbol{x}) \Big) d\boldsymbol{x} \\
&= \sum_{i=1}^n \int_{\mathbb{R}^{n-1}} \underbrace{\Big(\int_{\mathbb{R}} x_i \cdot \frac{\partial}{\partial x_i} f(\boldsymbol{x}) dx_i \Big)}_{=-\int_{\mathbb{R}} f(\boldsymbol{x}) dx_i} d\boldsymbol{x}_{-i} = -n \int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x} = -n
\end{aligned}$$

where $\boldsymbol{x}_{-i} \in \mathbb{R}^{n-1}$ is the vector $\boldsymbol{x}$ without the ith entry. □

We can now get a nice application out of this theorem. If we take $(n-1)$-dimensional slices of an arbitrary convex body, then the maximum volume slice might not go though the barycenter — but one can prove that no slice is more than a constant factor larger than the one through the barycenter.

Lemma 10.16. *Let $K \subseteq \mathbb{R}^n$ be a centered convex body with $\mathrm{Vol}_n(K) = 1$. Fix a direction $\boldsymbol{\theta} \in S^{n-1}$ and consider the function $f(t) := \mathrm{Vol}_{n-1}(K \cap \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle = t\})$. Then for all $t \in \mathbb{R}$ one has $f(t) \leq e \cdot f(0)$.*



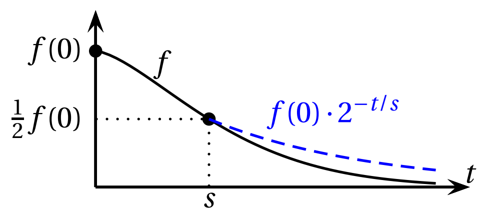
Proof. By *Brunn's Concavity Principle*, the function f is log-concave. Moreover, as K is centered, also f is centered. Then applying Theorem 10.15 with $n = 1$ gives $f(t) \leq e \cdot f(0)$ for all $t \in \mathbb{R}$. □

Now we can summarize a few useful properties of log concave distributions:

Theorem 10.17 (Properties of Log-Concave Distributions on $\mathbb{R}$)**.** *Let X be an $\mathbb{R}$-valued continuous random variable with $\mathbb{E}[X] = 0$ so that the density function $f : \mathbb{R} \to [0,\infty[$ is log-concave. Then the following holds for some universal constants $C_1, C_2 > 0$:*

(A) One has $\Pr[X \geq 0] \geq \frac{1}{e}$.
(B) One has $f(t) \leq 6f(0) \cdot 2^{-|t| \cdot f(0)/2}$ for any $t \in \mathbb{R}$.
(C) The standard deviation of X satisfies $\frac{C_1}{f(0)} \leq \mathbb{E}[X^2]^{1/2} \leq \frac{C_2}{f(0)}$.
(D) For any $\lambda \geq 0$ one has $\Pr[|X| \geq \lambda \cdot \frac{2}{f(0)}] \leq 40 \cdot 2^{-\lambda}$.

Proof. **Claim (A).** The claim is already shown in the proof of Grünbaum's Lemma (Lemma 1.38).
Claim (B). By symmetry it suffices to consider the case $t \geq 0$. Fix a value $s > 0$ so that $f(s) = \frac{1}{2}f(0)$ (after a small perturbation, f would be continuous and such a value exists). In particular $f(t) \geq \frac{1}{2}f(0)$ for $0 \leq t \leq s$ and hence $s \leq \frac{2}{f(0)}$ as $\int_{\mathbb{R}} f(t)dt = 1$. On the other hand, f is log concave and so for $t > s$ one has $\frac{1}{2}f(0) = f(s) \geq f(0)^{1-s/t} f(t)^{s/t}$ which can be rearranged to $f(t) \leq f(0) \cdot 2^{-t/s} \leq f(0) \cdot 2^{-t \cdot f(0)/2}$.



For $0 \leq t \leq s$ we use the estimate $f(t) \leq 3f(0)$ from Fradelizi's Theorem (Theorem 10.15). Both cases are dominated by the upper bound $f(t) \leq 6f(0) \cdot 2^{-t \cdot f(0)/2}$.
Claim (C). Again we use the upper bound $\|f\|_\infty \leq 3f(0)$ from Fradelizi's Theorem (Theorem 10.15). This implies that $\Pr[|X| \geq \frac{1}{12f(0)}] \geq 1 - \frac{2\|f\|_\infty}{12f(0)} \geq \frac{1}{2}$. Then we can lower bound the standard deviation as $\mathbb{E}[X^2]^{1/2} \geq (\frac{1}{2} \cdot \frac{1}{(12f(0))^2})^{1/2} \geq \frac{1}{24f(0)}$. For the upper bound we use the exponential decay from (B) to derive that

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} t^2 f(t)dt \overset{(B)}{\leq} 2 \cdot 6f(0) \int_0^{\infty} t^2 \cdot 2^{-t \cdot f(0)/2} dt = 12f(0) \cdot \frac{16}{\ln(2)^3} \cdot \frac{1}{f(0)^3} \leq \frac{600}{f(0)^2}$$

Claim (D). By symmetry it suffices to prove that $\Pr[X \geq \lambda \frac{2}{f(0)}] \leq 20 \cdot 2^{-\lambda}$. We abbreviate $s := \lambda \frac{2}{f(0)}$. Then

$$\Pr[X \geq t] = \int_s^{\infty} f(t)dt \overset{(B)}{\leq} 6f(0) \cdot \int_s^{\infty} 2^{-t \cdot f(0)/2} dt = \frac{2 \cdot 6}{\ln(2)} \cdot 2^{-sf(0)/2} \leq 20 \cdot 2^{-\lambda}$$

$\square$

10.7 Slices of isotropic bodies

Now we can prove that for a convex body in isotropic position, *every* slice though the origin has approximately the same volume of $\Theta(\frac{1}{L_K})$. Recall that we abbreviate $\boldsymbol{\theta}^{\perp} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \perp \boldsymbol{\theta}\}$ as the $(n-1)$-dimensional subspace orthogonal to $\boldsymbol{\theta}$.

Theorem 10.18. *Let $K \subseteq \mathbb{R}^n$ be an isotropic convex body. Then for every direction $\boldsymbol{\theta} \in S^{n-1}$ one has*

$$\frac{c_1}{L_K} \leq Vol_{n-1}(K \cap \boldsymbol{\theta}^{\perp}) \leq \frac{c_2}{L_K}$$

where $c_1, c_2 > 0$ are absolute constants.

Proof. We fix a body K in isotropic position and a direction $\boldsymbol{\theta}$ and abbreviate $f(t) := \mathrm{Vol}_{n-1}(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle = t\})$ as the function that gives the volume of the slices in direction $\boldsymbol{\theta}$. Recall that f is log concave. Consider the random variable $\langle \boldsymbol{x}, \boldsymbol{\theta} \rangle$ for $\boldsymbol{x} \sim K$ which has f as density function. Then

$$L_K = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim K}\left[\langle \boldsymbol{x}, \boldsymbol{\theta} \rangle^2\right]^{1/2} \stackrel{\text{Thm 10.17.(C)}}{\in} \left[\frac{C_1}{f(0)}, \frac{C_2}{f(0)}\right]$$

Rearranging gives the claim. ◻

This brings us to the following conjecture (also called the *Slicing Problem* or the *Hyperplane Conjecture*):

Conjecture 3 (Slicing Conjecture I)**.** *For every convex body $K \subseteq \mathbb{R}^n$ in isotropic position and any direction $\boldsymbol{\theta} \in S^{n-1}$ one has*

$$Vol_{n-1}(K \cap \boldsymbol{\theta}^{\perp}) \geq C_1$$

where $C_1 > 0$ is a universal constant.

Conjecture 4 (Slicing Conjecture II)**.** *For every convex body $K \subseteq \mathbb{R}^n$ with $Vol_n(K) = 1$ and $bary(K) = \mathbf{0}$, there exists a direction $\boldsymbol{\theta} \in S^{n-1}$ so that*

$$Vol_{n-1}(K \cap \boldsymbol{\theta}^{\perp}) \geq C_2$$

where $C_2 > 0$ is a universal constant.

We will now show these conjectures are equivalent:

Lemma 10.19. *Isotropic Constant Conjecture I ⇔ Slicing Conjecture I ⇔ Slicing Conjecture II.*

Proof. We prove the equivalences in several steps:

Claim. Isotropic Constant Conjecture I ⇔ Slicing Conjecture I.

Proof of Claim. From Theorem 10.18 we know that for a convex body K in isotropic position one has $\mathrm{Vol}_{n-1}(K \cap \boldsymbol{\theta}^{\perp}) = \Theta(\frac{1}{L_K})$ for every $\boldsymbol{\theta} \in S^{n-1}$. □

Claim. Slicing Conjecture II ⇒ Slicing Conjecture I.

Proof of Claim. Suppose K is in isotropic position. Then by Slicing Conjecture II there is *one* direction $\boldsymbol{\theta}$ with $\mathrm{Vol}_{n-1}(K \cap \boldsymbol{\theta}^{\perp}) \geq \Omega(1)$. Then by Theorem 10.18 we know that this holds for *every* slice. □

Claim. Isotropic Constant Conjecture I ⇒ Slicing Conjecture II.

Proof of Claim. Let K be any centered convex body with $\mathrm{Vol}_n(K) = 1$. Then by Lemma 10.14 there is a direction $\boldsymbol{\theta} \in S^{n-1}$ with $\mathbb{E}_{\boldsymbol{x}\sim K}[\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle^2] \leq L_K^2$. Again consider the function $f(t) := \mathrm{Vol}_{n-1}(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle = t\})$ which is the density function of the random variable $\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle$ where $\boldsymbol{x} \sim K$. Then

$$\frac{C_1}{f(0)} \overset{\text{Thm 10.17.}(C)}{\leq} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\left[\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle^2\right]^{1/2} \leq L_K$$

Rearranging gives $f(0) \geq \Omega(\frac{1}{L_K})$ and the claim follows. □

10.8 Bourgain's upper bound for the isotropic constant

The best bounds known on the quantity L_K are $O(n^{1/4}\ln(n))$ due to Bourgain and $O(n^{1/4})$ due to Klartag as well as a very recent bound of Chen [Che20] which shows that $L_K \leq n^{o(1)}$. We will prove the weaker bound of Bourgain; here in the exposition we will follow the simple proof of Dar. First, we need an estimate about the distribution of volume:

Lemma 10.20. *There is a constant $C > 0$ so that the following holds: Let $K \subseteq \mathbb{R}^n$ be an isotropic convex body. For a direction $\boldsymbol{\theta} \in S^{n-1}$ one has*

$$\Pr_{\boldsymbol{x}\sim K}\Big[|\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle| \geq \lambda \cdot C \cdot L_K\Big] \leq 40 \cdot 2^{-\lambda}$$

for all $\lambda \geq 0$.

Proof. For a fixed direction $\boldsymbol{\theta} \in S^{n-1}$ we can consider the random variable $\boldsymbol{X} := \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle$ with $\boldsymbol{x} \sim K$ and its log concave density function $f(t) := \mathrm{Vol}_{n-1}(\{\boldsymbol{x} \in K \mid \langle \boldsymbol{x}, \boldsymbol{\theta}\rangle = t\})$. By isotropy of K and Theorem 10.17.(C) we have $L_K = \mathbb{E}[X^2]^{1/2} \in [\frac{C_1}{f(0)}, \frac{C_2}{f(0)}]$. Then by Theorem 10.17.(D) we get

$$\Pr\Big[|X| \geq \lambda \frac{2}{C_1} L_K\Big] \leq \Pr\Big[|X| \geq \lambda \frac{2}{f(0)}\Big] \overset{\text{Thm 10.17.(D)}}{\leq} 40 \cdot 2^{-\lambda}$$

□

This can be understood that for any isotropic convex body and any direction $\boldsymbol{\theta}$, at least 99.9% of the volume lies within a slab of width $O(\frac{1}{L_K})$. We would like to point out that the concentration from Lemma 10.20 is tight, at least when one requires it to hold for every direction $\boldsymbol{\theta}$. To see this, consider $K := \Theta(n) \cdot B_1^n$ where the constant is chosen so that $\mathrm{Vol}_n(K) = 1$. Then $L_k = \Theta(1)$ as one can verify and the decay in any of the standard basis directions $\theta \in \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ decays as in the statement of the Lemma. That being said, for a *random direction* $\boldsymbol{\theta}$ the decay would be a lot stronger. There is a rather simple consequence of this lemma.

Corollary 10.21. *Let $K \subseteq \mathbb{R}^n$ be an isotropic convex body and let $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N \in S^{n-1}$ be directions. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\max_{i=1,\dots,N} |\langle \boldsymbol{\theta}_i, \boldsymbol{x}\rangle|\Big] \leq O(L_K \ln(N)).$$

Proof. For each $i \in [N]$ and $k \geq 1$ one has $\Pr[|\langle \boldsymbol{\theta}_i, \boldsymbol{x}\rangle| \geq kC' \ln(N)] \leq \frac{1}{N^k}$ for some constant $C' > 0$. Then apply union bound and sum up contributions for $k \geq 1$. The calculation is very similar to Lemma 3.17.(i) and we omit the details. □

We need a decomposition argument that is similar to the one from Dudley's Theorem from Chapter 9.

Theorem 10.22 (Dudley-Fernique Decomposition)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body with $\mathbf{0} \in K$ and let $r > 0$ be a parameter. Let m be so that $K \subseteq r2^m \cdot B_2^n$. Then for all $j \in \{1, \dots, m\}$ there are sets $Z_j \subseteq 3r \cdot 2^j \cdot B_2^n$ of cardinality*

$$|Z_j| \leq \exp\Big(O(n) \cdot \Big(\frac{w(K)}{r2^j}\Big)^2\Big)$$

so that the following holds: for every $\boldsymbol{x} \in K$, there is a decomposition $\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{z}_1 + \dots + \boldsymbol{z}_m$ where $\boldsymbol{z}_j \in Z_j$ and $\|\boldsymbol{w}\|_2 \leq 2r$.

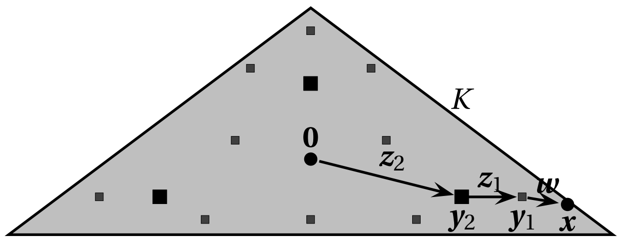
Proof. By *Sudakov's Theorem* (Theorem 4.12), we can cover K with balls of radius $2^j r$ where the centers $N_j \subseteq K$ satisfy $|N_j| \leq \exp(O(n) \cdot (\frac{w(K)}{2^j r})^2)$. Here one can choose $N_m = \{\mathbf{0}\}$. For $j \geq 1$, we define

$$Z_j := (N_j - N_{j+1}) \cap (3r2^j B_2^n)$$

as the set of difference vectors between step j and $j-1$ that are short enough. After adapting the constant, we still have $|Z_j| \leq \exp(O(n) \cdot (\frac{1}{r2^j} \cdot w(K))^2)$. Now fix

$\boldsymbol{x} \in K$. Let $\boldsymbol{y}_j \in N_j$ be the point with $\|\boldsymbol{x}-\boldsymbol{y}_j\|_2 \leq r \cdot 2^j$. Setting $\boldsymbol{z}_j := \boldsymbol{y}_j - \boldsymbol{y}_{j+1}$ where $\boldsymbol{y}_{m+1} := \boldsymbol{0}$ and $\boldsymbol{w} := \boldsymbol{x} - \boldsymbol{y}_1$ gives a decomposition of

$$\boldsymbol{x} = \underbrace{(\boldsymbol{x}-\boldsymbol{y}_1)}_{=\boldsymbol{w}} + \underbrace{(\boldsymbol{y}_1-\boldsymbol{y}_2)}_{=\boldsymbol{z}_1} + \underbrace{(\boldsymbol{y}_2-\boldsymbol{y}_3)}_{=\boldsymbol{z}_2} + \ldots + \underbrace{(\boldsymbol{y}_m-\boldsymbol{y}_{m+1})}_{=\boldsymbol{z}_m}$$



Visualization for $m = 2$

Note that indeed $\|\boldsymbol{w}\|_2 \leq 2r$ and $\|\boldsymbol{z}_j\|_2 \leq \|\boldsymbol{x}-\boldsymbol{y}_j\|_2 + \|\boldsymbol{x}-\boldsymbol{y}_{j+1}\|_2 \leq r \cdot (2^j + 2^{j+1}) \leq 3r \cdot 2^j$ so that $\boldsymbol{z}_j \in Z_j$. $\square$

Now we can give the result of Bourgain, where we use the proof of Dar:

Theorem 10.23 (Bourgain)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then $L_K \leq O(n^{1/4}\ln(n))$.*

Proof. Wl.o.g. assume that K is in isotropic position. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the volume preserving linear map from the *$\ell\ell^\circ$-estimate* from Chapter 6, which means that $w(T(K)) \leq O(\sqrt{n}\log(n))$ and $\mathrm{Vol}_n(T(K)) = 1$. Recall that the underlying matrix $\boldsymbol{T}$ can be chosen to be symmetric and positive definite. As K itself is in isotropic position we have $\mathbb{E}_{\boldsymbol{x}\sim K}[\boldsymbol{x}\boldsymbol{x}^T] = L_K^2 \boldsymbol{I}_n$. Then

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\max_{\boldsymbol{y}\in T(K)} \langle \boldsymbol{x}, \boldsymbol{y}\rangle\Big] \geq \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}[\langle \boldsymbol{x}, \boldsymbol{T}\boldsymbol{x}\rangle] = \langle \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}[\boldsymbol{x}\boldsymbol{x}^T]}_{=L_K^2 \boldsymbol{I}_n}, \boldsymbol{T}\rangle = L_K^2 \cdot \mathrm{Tr}[\boldsymbol{T}] \geq L_K^2 n \quad (*)$$

where we use that $\det(\boldsymbol{T}) = 1 \Rightarrow \mathrm{Tr}[\boldsymbol{T}] \geq n$.

We can apply the *Dudley-Fernique Decomposition* from Theorem 10.22 to the body $T(K)$ with a parameter $r > 0$ that we determine later. We will use that every vector $\boldsymbol{y} \in T(K)$ can be decomposed into $\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{z}_1 + \ldots + \boldsymbol{z}_m$ where $\|\boldsymbol{w}\|_2 \leq 2r$ and $\boldsymbol{z}_j \in Z_j$ with $\|\boldsymbol{z}_j\|_2 \leq 3r \cdot 2^j$ and $|Z_j| \leq \exp(O(n) \cdot (\frac{w(T(K))}{r2^j})^2)$. For a vector $\boldsymbol{z}$, we abbreviate $\bar{\boldsymbol{z}} := \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_2}$ as his normalization. Then we can continue the bound from

$(*)$ to

$$\begin{aligned}
L_K^2 n &\le \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\max_{\boldsymbol{y}\in T(K)}\langle \boldsymbol{x},\boldsymbol{y}\rangle\Big] \le \sum_{j=1}^m \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\max_{\boldsymbol{z}\in Z_j}\langle \boldsymbol{z},\boldsymbol{x}\rangle\Big] + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\max_{\|\boldsymbol{w}\|_2\le 2r}\langle \boldsymbol{w},\boldsymbol{x}\rangle\Big] \\
&\le \sum_{j=1}^m 3r\cdot 2^j \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}\Big[\max_{\boldsymbol{z}\in Z_j}\langle \bar{\boldsymbol{z}},\boldsymbol{x}\rangle\Big]}_{\text{bound by Cor 10.21}} + 2r\cdot \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim K}[\|\boldsymbol{x}\|_2]}_{\le\sqrt{n}L_K \text{ by Lem 10.3}} \\
&\le \sum_{j=1}^m 3r\cdot 2^j\cdot O(L_K\ln(|Z_j|)) + 2r\sqrt{n}L_K \\
&\le \sum_{j=1}^m 3r2^jL_K\cdot O(n)\cdot\Big(\frac{w(T(K))}{r2^j}\Big)^2 + 2r\sqrt{n}L_K \\
&\le \frac{O(n^2\log^2(n))}{r}L_K\underbrace{\sum_{j=1}^m 2^{-j}}_{\le 1} + 2r\sqrt{n}L_K
\end{aligned}$$

We divide the obtained relation by $L_K n$ and choose $r := n^{3/4}\log(n)$ to balance the terms and get

$$L_K \le \frac{O(n\log^2(n))}{r} + \frac{2r}{\sqrt{n}} \le O(n^{1/4}\log(n))$$

□

Gaussian tails. It is also possible to obtain concentration with Gaussian-type tails for volumes in isotropic bodies, not just the exponential tails as in Lemma 10.20. We state the result without proof:

Lemma 10.24. *For any convex body $K\subseteq\mathbb{R}^n$ in isotropic position one has*

$$Vol_n\big(\{\boldsymbol{x}\in K \mid \|\boldsymbol{x}\|_2 \ge C\sqrt{n}L_K\cdot\lambda\}\big) \le 2e^{-\lambda^2}$$

for any $\lambda\ge 0$, where $C>0$ is a universal constant.

Chapter 11

Projections and sections of cubes

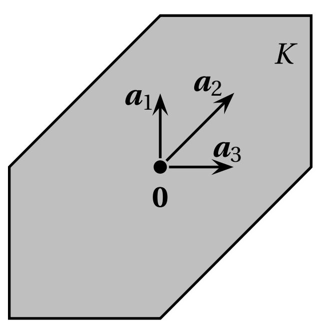
We have met cubes B_∞^n as fundamental, yet very simple convex bodies. It turns out that projections and sections of cubes have interesting properties that we want to discuss in this chapter.

11.1 Introduction to zonotopes

Given vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m \in \mathbb{R}^n$, the set

$$K := \mathrm{conv}\{-\boldsymbol{a}_1,+\boldsymbol{a}_1\}+\ldots+\mathrm{conv}\{-\boldsymbol{a}_m,+\boldsymbol{a}_m\} = \Big\{\sum_{i=1}^m \lambda_i \boldsymbol{a}_i : \lambda_1,\dots,\lambda_m \in [-1,1]\Big\}$$

is called a *zonotope*. The 1-dimensional parts $\mathrm{conv}\{-\boldsymbol{a}_i,+\boldsymbol{a}_i\}$ are called the *segments* defining the zonotope. In our definition, a zonotope is always a compact symmetric convex set. The easiest zonotope is certainly the cube B_∞^n itself. Here is a zonotope in $\mathbb{R}^2$ defined by 3 segments:



More generally, a *zonoid* is any body that can be arbitrarily well approximated by zonotopes. Formally speaking a symmetric convex body $K \subseteq \mathbb{R}^n$ is a zonoid if there is a sequence $\{K_t\}_{t\in\mathbb{N}}$ of zonotopes so that $\lim_{t\to\infty} d_H(K,K_t) = 0$ where

d_H is the *Hausdorff metric* (see Chapter 1.3). It turns out that for all $p \geq 2$, the balls B_p^n are zonoids while for $1 \leq p < 2$, the balls B_p^n are not zonoids. For more fascinating details on zonotopes and zonoids we refer to the article of Bourgain, Lindenstrauss and Milman [BLM89].

Recall that the support function $h_K : \mathbb{R}^n \to \mathbb{R}$ of a convex body K is $h_K(\boldsymbol{a}) := \sup_{\boldsymbol{x}\in K} \langle \boldsymbol{a}, \boldsymbol{x}\rangle$. We will see that the support function of zonotopes has a neat expression. For this it will be notationally convinient to think of the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ as the rows of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$.

Lemma 11.1. *Consider the zonotope $K = \{\sum_{i=1}^m y_i \boldsymbol{A}_i \mid \boldsymbol{y} \in [-1,1]^m\}$ where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Then the support function is $h_K(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|_1$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Proof. We verify that

$$h_K(\boldsymbol{x}) = \max\Big\{\sum_{i=1}^m y_i \langle \boldsymbol{A}_i, \boldsymbol{x}\rangle : y_i \in [-1,1]\Big\} = \sum_{i=1}^m |\langle \boldsymbol{A}_i, \boldsymbol{x}\rangle| = \|\boldsymbol{A}\boldsymbol{x}\|_1$$

□

If we consider the linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ with $T(\boldsymbol{y}) = \boldsymbol{A}^T\boldsymbol{y}$ then the zonotope $K = \{\sum_{i=1}^m y_i \boldsymbol{A}_i \mid \boldsymbol{y} \in [-1,1]^m\}$ is of the form $K = T(B_\infty^m)$, which means that one can think of zonotopes as *projections of cubes*. Note that here "projection" does not necessarily mean "orthogonal projection".

We want to at least prove that B_2^n is a zonoid. In fact, we will even prove a rather strong quantitative bound on the number of segments required for an ε-approximation. This will also provide us with some motivation for the topic of the following section.

Theorem 11.2. *Let $0 < \varepsilon \leq \frac{1}{2}$. Then there is a zonotope $K \subseteq \mathbb{R}^n$ with $O(\frac{n}{\varepsilon^2})$ many segments so that $(1-\varepsilon)\cdot K \subseteq B_2^n \subseteq (1+\varepsilon)\cdot K$.*

Proof. This will be a quick application of the incredibly powerful and flexible concentration inequality of Theorem 9.29. Draw a Gaussian random matrix $\boldsymbol{A} \sim N^{m\times n}(0,1)$. We verify that for each $\boldsymbol{x} \in S^{n-1}$ we have $\mathbb{E}_{\boldsymbol{A}}[\|\boldsymbol{A}\boldsymbol{x}\|_1] = m\mathbb{E}_{\boldsymbol{a}\sim N(\boldsymbol{0},\boldsymbol{I}_n)}[|\langle \boldsymbol{a}, \boldsymbol{x}\rangle|] = m\sqrt{\frac{2}{\pi}} =: \mu$. Note that $\frac{1}{\sqrt{m}}B_2^m \subseteq B_1^m$. Then

$$\mathop{\mathbb{E}}_{\boldsymbol{A}}\Big[\sup_{\boldsymbol{x}\in S^{n-1}} \big|\|\boldsymbol{A}\boldsymbol{x}\|_1 - \mu\big|\Big] \overset{\text{Theorem 9.29}}{\leq} O(\sqrt{m})\cdot \underbrace{g(S^{n-1})}_{\leq O(\sqrt{n})} \leq O(\sqrt{nm}) \leq \frac{\varepsilon}{2}\cdot\mu$$

where in the last step we use $m \geq C\frac{n}{\varepsilon^2}$ for a large enough constant $C > 0$. Then if K is the zonotope generated by the m vectors $\frac{\boldsymbol{A}_1}{\mu}, \ldots, \frac{\boldsymbol{A}_m}{\mu}$ then the support function satisfies $|h_K(\boldsymbol{x}) - 1| \leq \frac{\varepsilon}{2}$ for all $\boldsymbol{x} \in S^{n-1}$ and so K satisfies the claim. □

Note that for a weaker bound of $O(\log(\frac{1}{\varepsilon})\cdot\frac{n}{\varepsilon^2})$ it suffices to use an ε-net argument. Considering that a ball B_2^n can be approximated with a zonotope with only linear in n many segments, we might be wondering how well an *arbitrary zonoid* can be approximated. We will answer this next.

11.2 Approximating zonotopes with few segments

In this section we will prove that every zonoid can be approximated surprisingly well with only a few segments. We provide the proof of the best known bound which is due to Talagrand:

Theorem 11.3 (Talagrand 1990)**.** *Let $0<\varepsilon\le\frac{1}{2}$. For any zonoid $K\subseteq\mathbb{R}^n$ there is a zonotope Q with $(1-\varepsilon)Q\subseteq K\subseteq(1+\varepsilon)Q$ so that Q has at most $O(\frac{n\log(n)}{\varepsilon^2})$ many segments.*

It is not known whether the $\log(n)$ term can be removed (even if one would be willing to pay with a higher dependence on ε). By definition we can approximate the initial zonoid arbitrarily well with a zonotope $\tilde{K}=\{\sum_{i=1}^m y_i\boldsymbol{A}_i \mid \boldsymbol{y}\in[-1,1]^m\}$. Then the support function of that zonotope is $h_{\tilde{K}}(\boldsymbol{x})=\|\boldsymbol{A}\boldsymbol{x}\|_1$ and our goal is to approximate that support function with $h_Q(\boldsymbol{x})=\|\boldsymbol{B}\boldsymbol{x}\|_1$ where the number of rows of $\boldsymbol{B}$ is at most $O(\frac{n\log(n)}{\varepsilon^2})$ (as that number corresponds to the number of segments of Q). We will do exactly that in an iterative process.

11.2.1 Reducing the number of segments by a constant factor

Before we come to the main technical lemma, we need to review some more linear algebra. Consider a matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ with $r:=\mathrm{rank}(\boldsymbol{A})$ and let $\boldsymbol{A}=\sum_{k=1}^r\sigma_k\boldsymbol{u}_k\boldsymbol{v}_k^T$ be the *Singular Value Decomposition* where $\boldsymbol{u}_1,\dots,\boldsymbol{u}_r\in\mathbb{R}^m$ and $\boldsymbol{v}_1,\dots,\boldsymbol{v}_r\in\mathbb{R}^n$ are both orthonormal. Then the *Frobenius norm* is $\|\boldsymbol{A}\|_F=(\sum_{j=1}^n\|\boldsymbol{A}^j\|_2^2)^{1/2}=(\sum_{i=1}^m\|\boldsymbol{A}_i\|_2^2)^{1/2}$.

We define the *pseudo-inverse* of $\boldsymbol{A}$ as $\boldsymbol{A}^+:=\sum_{k=1}^r\frac{1}{\sigma_k}\boldsymbol{v}_k\boldsymbol{u}_k^T\in\mathbb{R}^{n\times m}$. Now, let us assume that the matrix has *full column rank*, i.e. $\mathrm{rank}(\boldsymbol{A})=n$ (which means in particular that $m\ge n$). Then it is worth noting that

$$\boldsymbol{A}^+\boldsymbol{A}=\sum_{k=1}^n\sum_{\ell=1}^n\frac{\sigma_k}{\sigma_\ell}\boldsymbol{v}_\ell\boldsymbol{u}_\ell^T\boldsymbol{u}_k\boldsymbol{v}_k^T=\sum_{k=1}^n\boldsymbol{v}_k\boldsymbol{v}_k^T=\boldsymbol{I}_n$$

using orthonormality. On the other hand, $\boldsymbol{A}\boldsymbol{A}^+=\sum_{k=1}^n\boldsymbol{u}_k\boldsymbol{u}_k^T$ is a symmetric $m\times m$ matrix with n Eigenvalues that are 1 and all other Eigenvalues 0. In particular the column lengths satisfy $\sum_{i=1}^{m} \|(\boldsymbol{A}\boldsymbol{A}^{+})^{i}\|_2^2 = \|\boldsymbol{A}\boldsymbol{A}^{+}\|_F^2 = n$. Hence on average the columns of $\boldsymbol{A}\boldsymbol{A}^{+}$ have length $\mathbb{E}_{i\sim[m]}[\|(\boldsymbol{A}\boldsymbol{A}^{+})^{i}\|_2^2] = \frac{n}{m}$. Now we come to the main technical lemma where we show how to reduce the number of rows by a constant fraction while approximately preserving the norm $\|\boldsymbol{A}\boldsymbol{x}\|_1$ for all $\boldsymbol{x}$.

Lemma 11.4. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be a matrix with rank$(\boldsymbol{A}) = n$. Then there is a matrix $\boldsymbol{B} \in \mathbb{R}^{\frac{3}{4}m\times n}$ so that*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^n} \frac{|\|\boldsymbol{B}\boldsymbol{x}\|_1 - \|\boldsymbol{A}\boldsymbol{x}\|_1|}{\|\boldsymbol{A}\boldsymbol{x}\|_1} \leq O\Big(\sqrt{\frac{n\log(m)}{m}}\Big)$$

The overall proof strategy is simple: for every row index i flip a fair coin and with probability $\frac{1}{2}$ we double the row and with probability $\frac{1}{2}$ we replace it with $\mathbf{0}$. The first problem is that some rows might be more important than others, for example if some row $\boldsymbol{A}_i$ happens to be orthogonal to all others, then in order to have a relative error < 1 we definitely need to keep that row. We solve that problem by simply fixing the rows that are long in carefully defined sense. Another problem is that the absolute error that we allow for a vector $\boldsymbol{x}$ actually depends on the quantity $\|\boldsymbol{A}\boldsymbol{x}\|_1$ which is inconvinient.

Proof. See Talagrand. □

11.2.2 Finishing the proof of Talagrand's bound

As discussed earlier to obtain Talagrand's result from Theorem 11.3 it suffices to prove the following:

Lemma 11.5. *Let $0 < \varepsilon \leq \frac{1}{2}$ and $n \in \mathbb{N}$. Then for any matrix $\boldsymbol{A} \in \mathbb{R}^{m_0\times n}$ with rank$(\boldsymbol{A}) = n$ there is a matrix $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ so that $m \leq O(\frac{n\log n}{\varepsilon^2})$ and*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^n} \frac{|\|\boldsymbol{B}\boldsymbol{x}\|_1 - \|\boldsymbol{A}\boldsymbol{x}\|_1|}{\|\boldsymbol{A}\boldsymbol{x}\|_1} \leq \varepsilon$$

Proof. Let $C_1 > 0$ be the implicit constant from the claim of Lemma 11.4. We abbreviate the initial matrix by $\boldsymbol{A}^{(0)} := \boldsymbol{A} \in \mathbb{R}^{m_0\times n}$. More generally for $t \geq 0$ we will have a matrix of the form $\boldsymbol{A}^{(t)} \in \mathbb{R}^{m_t\times n}$. As long as we have $m_t > \frac{Dn\log(n)}{\varepsilon^2}$ for a large enough constant $D > 0$, we apply Lemma 11.4 and find a matrix $\boldsymbol{A}^{(t+1)} \in \mathbb{R}^{m_{t+1}\times n}$ with $m_{t+1} \leq \frac{3}{4}m_t$ rows so that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^n} \frac{|\|\boldsymbol{A}^{t+1}\boldsymbol{x}\|_1 - \|\boldsymbol{A}^{(t)}\boldsymbol{x}\|_1|}{\|\boldsymbol{A}^{(t)}\boldsymbol{x}\|_1} \leq C_1\sqrt{\frac{n\log(m_t)}{m_t}}$$

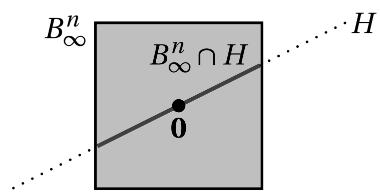
Let T be the first index that satisfies $m_T \leq \frac{Dn\log(n)}{\varepsilon^2}$. Then for any fixed $\boldsymbol{x} \in \mathbb{R}^n \setminus \{0\}$ we have

$$\frac{\|\boldsymbol{A}^{(T)}\boldsymbol{x}\|_1}{\|\boldsymbol{A}^{(0)}\boldsymbol{x}\|_1} \leq \prod_{t=0}^{T-1}\left(1+C_1\sqrt{\frac{n\log(m_t)}{m_t}}\right) \overset{(*)}{\leq} 1+C_2\sqrt{\frac{n\log(m_{T-1})}{m_{T-1}}} \leq 1+\varepsilon$$

as $m_{T-1} \geq \frac{Dn\log(n)}{\varepsilon^2}$ and D is large enough. Note that the crucial point for $(*)$ is that the terms $\sqrt{\frac{n\log(m_t)}{m_t}}$ are geometrically increasing and hence are dominated by the last term. The lower bound on $\frac{\|\boldsymbol{A}^{(T)}\boldsymbol{x}\|_1}{\|\boldsymbol{A}^{(0)}\boldsymbol{x}\|_1}$ is analogous. Then $\boldsymbol{B} := \boldsymbol{A}^{(T)}$ satisfies the claim. $\square$

11.3 Vaaler's Theorem

Now we switch topics and talk about *sections* of cubes rather than *projections*. Consider the cube B_∞^n and a d-dimensional subspace $H \subseteq \mathbb{R}^n$. We are wondering what lower bound one can prove on $\mathrm{Vol}_d(B_\infty^n \cap H)$.



If H happens to be spanned by d coordinate directions $\boldsymbol{e}_i$, then clearly $\mathrm{Vol}_d(B_\infty^n \cap H) = \mathrm{Vol}_d(B_\infty^d) = 2^d$. A beautiful theorem of Vaaler tells us that this indeed is the minimal volume for any subspace H. We will prove this result in the form of $\mathrm{Vol}_d([-\frac{1}{2}, \frac{1}{2}]^n \cap H) \geq 1$ which is notationally more convinient.

11.3.1 Comparision of distributions and Kanter's Lemma

A crucial ingredient in the proof of Vaaler's Theorem will be the following notion:

Definition 11.6. Let ν and μ be distributions on $\mathbb{R}^n$. We say that μ *is more peaked than* ν if

$$\nu(K) \leq \mu(K) \quad \forall \text{symmetric, closed, convex } K \subseteq \mathbb{R}^n$$

We write $\nu \preceq_{\text{peaked}} \mu$.

Intuitively, if $\nu \preceq_{\text{peaked}} \mu$ then the mass of μ is closer to the origin.

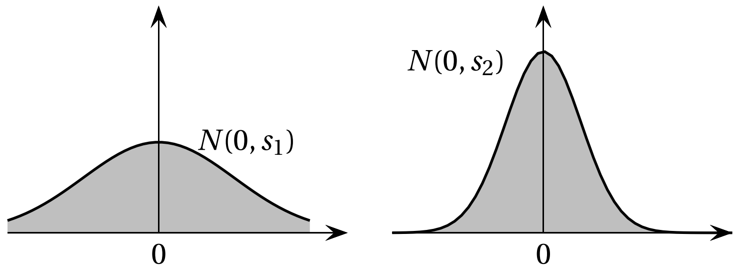


Fig: Two 1-dim Gaussians where $N(0,s_1) \preceq_{\text{peaked}} N(0,s_2)$ for $s_1 \geq s_2$.

In the following, for a compact set $Q \subseteq \mathbb{R}^n$, let Uniform(Q) be the uniform distribution on Q. Recall that for $s > 0$, $N(0,s^2)$ is the 1-dimensional Gaussian distribution with density function $\frac{1}{s\sqrt{2\pi}} \cdot \exp(-\frac{x^2}{2s^2})$.

We state a few convinient facts without proof (we will not even need all those facts for our purpose):

Lemma 11.7. *The following holds*

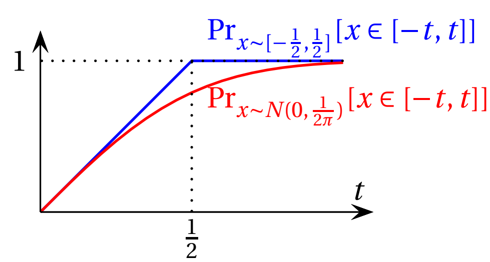
(i) If $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ are matrices with $0 \preceq \boldsymbol{A} \preceq \boldsymbol{B}$ then $N(\boldsymbol{0}, \boldsymbol{B}) \preceq_{peaked} N(\boldsymbol{0}, \boldsymbol{A})$.

(ii) For two symmetric convex bodies $K \subseteq Q$ one has Uniform$(Q) \preceq_{peaked}$ Uniform(K).

(iii) One has $N(0, \frac{1}{2\pi}) \preceq_{peaked}$ Uniform$([-\frac{1}{2}, \frac{1}{2}])$.

(iv) For any $s > 0$ one has $N(0, \frac{2}{\pi}s^2) \preceq_{peaked}$ Uniform$([-s,s])$.

For (iii) note that the density functions of $N(0, \frac{1}{2\pi})$ and Uniform$([-\frac{1}{2}, \frac{1}{2}])$ both have value 1 at the origin, hence the constants we have chosen are tight.



We will crucially rely on the following lemma due to Kanter:

Lemma 11.8 (Kanter 1977)**.** *Let ν_1, μ_1 be log-concave distributions on $\mathbb{R}^{n_1}$ with $\nu_1 \preceq_{peaked} \mu_1$ and let ν_2, μ_2 be log-concave distributions on $\mathbb{R}^{n_2}$ with $\nu_2 \preceq_{peaked} \mu_2$. Then the product distributions $\nu_1 \otimes \nu_2$ and $\mu_1 \otimes \mu_2$ are log-concave distributions on $\mathbb{R}^{n_1+n_2}$ and $(\nu_1 \otimes \nu_2) \preceq_{peaked} (\mu_1 \otimes \mu_2)$.*

A proof can be found for example in Chapter 4 of [JL01]. Note that in particular Uniform(K) is log-concave for any convex body K and for $\boldsymbol{A} \succeq 0$, the Gaussian distribution $N(\mathbf{0}, \boldsymbol{A})$ is log-concave.

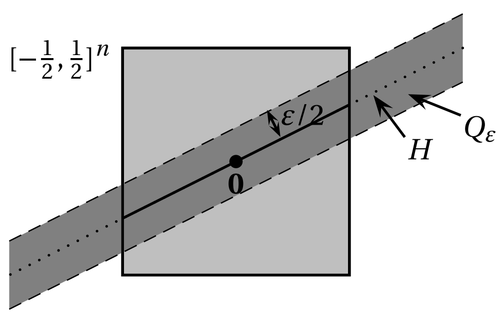
11.3.2 Proof of Vaaler's Theorem

Now, we are ready for the main proof.

Theorem 11.9 (Vaaler 1979)**.** *For any subspace $H \subseteq \mathbb{R}^n$ with $d := \dim(H)$ one has $Vol_d([-\frac{1}{2}, \frac{1}{2}]^n \cap H) \geq 1$.*

Proof. Our strategy is use the notion of peakedness and compare the volume with a corresponding Gaussian measure. First, recall that $\frac{1}{\sqrt{2\pi}} \cdot N(0,1) \preceq_{\text{peaked}}$ Uniform($[-\frac{1}{2}, \frac{1}{2}]$) by Lemma 11.7.(iii) and so by Kanter's Lemma this also holds for the n-fold product measure, i.e. $\frac{1}{\sqrt{2\pi}} \cdot N(\mathbf{0}, \boldsymbol{I}_n) \preceq_{\text{peaked}}$ Uniform($[-\frac{1}{2}, \frac{1}{2}]^n$). It will be convinient to abbreviate $\nu := \frac{1}{\sqrt{2\pi}} \cdot N(\mathbf{0}, \boldsymbol{I}_n)$ and $\mu :=$ Uniform($[-\frac{1}{2}, \frac{1}{2}]^n$). Recall that by a slight abuse of notation, the density function of ν is $\nu(\boldsymbol{x}) = \exp(-\pi\|\boldsymbol{x}\|_2^2)$. Next, let $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ be an orthonormal basis of $\mathbb{R}^n$ so that the first d columns are also a basis of H, i.e. $H = \text{span}\{\boldsymbol{B}^1, \ldots, \boldsymbol{B}^d\}$. Take a tiny value of $\varepsilon > 0$ and consider the intersection of strips

$$Q_\varepsilon := \left\{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{A}^j, \boldsymbol{x}\rangle| \leq \frac{\varepsilon}{2} \ \forall j = d+1, \ldots, n\right\}$$



Then using that Q_ε is a symmetric closed convex set we have

$$\varepsilon^{n-d} \text{Vol}_d(B_\infty^m \cap H) \approx \mu(Q_\varepsilon) \overset{\nu \preceq_{\text{peaked}} \mu}{\geq} \nu(Q_\varepsilon) = \Pr_{\boldsymbol{y} \sim \frac{1}{\sqrt{2\pi}} N(\mathbf{0}, \boldsymbol{I}_{n-d})}\left[\|\boldsymbol{y}\|_\infty \leq \frac{\varepsilon}{2}\right] \approx \varepsilon^{n-d} \underbrace{\nu(\mathbf{0})}_{=1}$$

Here the multiplicative error in both approximations "≈" goes to 0 as $\varepsilon \to 0$. Then rearranging and sending $\varepsilon \to 0$ gives the desired inequality of $\text{Vol}_d([-\frac{1}{2}, \frac{1}{2}]^n \cap H) \geq 1$. $\square$

Bibliography

[AAGM15] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[AS16] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[Bal97] Keith Ball. An elementary introduction to modern convex geometry. In *Flavors of geometry*, volume 31 of *Math. Sci. Res. Inst. Publ.*, pages 1–58. Cambridge Univ. Press, Cambridge, 1997.

[BLM89] Jean Bourgain, Joram Lindenstrauss, and Vitali Milman. Approximation of zonoids by zonotopes. *Acta mathematica*, 162(1):73–141, 1989.

[BM87] J. Bourgain and V. D. Milman. New volume ratio properties for convex symmetric bodies in $\mathbf{R}^n$. *Invent. Math.*, 88(2):319–340, 1987.

[Che20] Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the kls conjecture, 2020.

[Dvo59] A Dvoretzky. A theorem on convex bodies and applications to banach spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 45(2):223–226, 02 1959.

[Dvo61] Aryeh Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[FLM77] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139:53–94, 1977.

[FTJ79] T. Figiel and Nicole Tomczak-Jaegermann. Projections onto hilbertian subspaces of banach spaces. *Israel Journal of Mathematics*, 33(2):155–171, 1979.

[Gor85] Yehoram Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50(4):265–289, 1985.

[Gor88] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, Berlin, 1988.

[IRR+20] Siddharth Iyer, Anup Rao, Victor Reis, Thomas Rothvoss, and Amir Yehudayoff. An elementary exposition of pisier's inequality, 2020.

[JL01] William B. Johnson and Joram Lindenstrauss. Handbook of geometry of banach spaces. 2001.

[Joh48] Fritz John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[Kas77] B. Kashin. Section of some finite-dimensional sets and classes of smooth functions. 1977.

[KM87] Hermann König and Vitali D. Milman. On the covering numbers of convex bodies. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 82–95. Springer, Berlin, 1987.

[KM05] B. KLARTAG and V. MILMAN. Rapid steiner symmetrization of most of a convex body and the slicing problem. *Combinatorics, Probability and Computing*, 14(5-6):829–843, 2005.

[Lew79] D. R. Lewis. Ellipsoids defined by banach ideal norms. *Mathematika*, 26(1):18–29, 1979.

[Mat02] Jirí Matousek. *Lectures on discrete geometry*, volume 212 of *Graduate texts in mathematics*. Springer, 2002.

[Mil71] V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.

[Mil85] V. D. Milman. Random subspaces of proportional dimension of finite-dimensional normed spaces: approach through the isoperimetric inequality. In *Banach spaces (Columbia, Mo., 1984)*, volume 1166 of *Lecture Notes in Math.*, pages 106–115. Springer, Berlin, 1985.

[Mil86] Vitali D. Milman. Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(1):25–28, 1986.

[Mil88] V. D. Milman. Isomorphic symmetrization and geometric inequalities. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 107–131. Springer, Berlin, 1988.

[Mil90a] V. Milman. A note on a low M^*-estimate. In *Geometry of Banach spaces (Strobl, 1989)*, volume 158 of *London Math. Soc. Lecture Note Ser.*, pages 219–229. Cambridge Univ. Press, Cambridge, 1990.

[Mil90b] V. Milman. Spectrum of a position of a convex body and linear duality relations. In *Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part II (Ramat Aviv, 1989)*, volume 3 of *Israel Math. Conf. Proc.*, pages 151–161. Weizmann, Jerusalem, 1990.

[MP76] Bernard Maurey and Gilles Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.*, 58(1):45–90, 1976.

[Pis89a] Gilles Pisier. A new approach to several results of v. milman. *Journal fÃ¼r die reine und angewandte Mathematik*, 393:115–131, 1989.

[Pis89b] Gilles Pisier. *The Volume of Convex Bodies and Banach Space Geometry*. Cambridge Tracts in Mathematics. Cambridge University Press, 1989.

[PTJ86] Alain Pajor and Nicole Tomczak-Jaegermann. Subspaces of small codimension of finite-dimensional Banach spaces. *Proc. Amer. Math. Soc.*, 97(4):637–642, 1986.

[PVZ17] Grigoris Paouris, Petros Valettas, and Joel Zinn. Random version of dvoretzky's theorem in lpn. *Stochastic Processes and their Applications*, 127(10):3187 – 3227, 2017.

[STJ80] S. Szarek and N. Tomczak-Jaegermann. On nearly euclidean decomposition for some classes of banach spaces. *Compositio Mathematica*, 40(3):367–385, 1980.

[Sza77] S.J. Szarek. *On Kashin's Almost Euclidean Orthogonal Decomposition of L.* Polish Academy of Sciences [PAS]. Institute of Mathematics, 1977.

[Tal87] Michel Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.

[Ver19] Roman Vershynin. High-dimensional probability. 2019.

[Vri23] Beatrice-Helen Vritsiou. Regular ellipsoids and a blaschke-santaló-type inequality for projections of non-symmetric convex bodies, 2023.