I. Introduction

What is "Asymptotic Geometric Analysis"?

functional analysis

asymp. geom. analysis /
local Banach space theory /
asymp. geometric analysis /
Banach space geometry

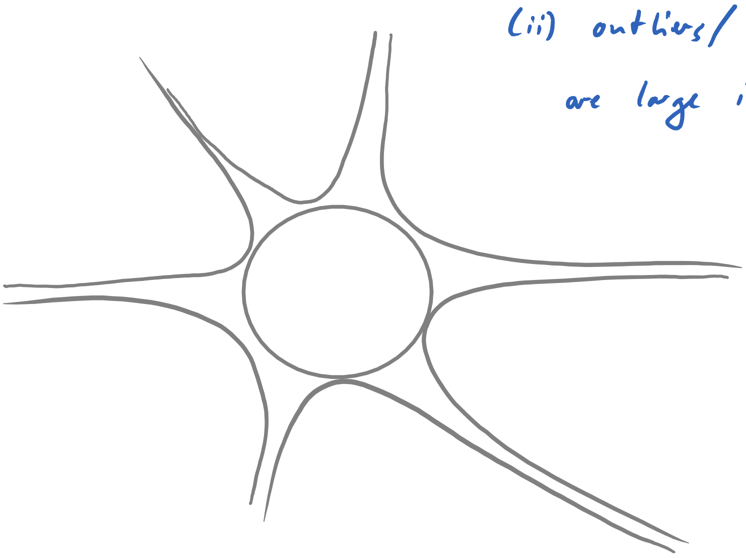
convex geometry

<u>typ. question</u>: how does this n-dim linear/convex structure look like, when n becomes very large?

<u>key tools</u>:
- probabilistic method
- concentration of measure

<u>applications</u>: dimensionality reduction in the analysis & processing of large data

<u>typ. picture</u>: a sometimes useful way to depict a convex body in high dimensions is to think of
(i) a bulk that contains almost all volume and is not too different from a Euclidean ball and
(ii) outliers/tentacles that contain little volume but are large in diameter.



Milman's "hyperbolic drawing" of a high dim. convex body.

This picture is helpful to understand phenomena like this:

Define $B_1^n := \{ x \in \mathbb{R}^n \mid \sum_i |x_i| \leq 1 \}$ and $K := \sqrt{n} B_1^n$.

The largest Euclidean Ball contained in K has radius 1 although K has diameter $2\sqrt{n}$. If U is a random orthogonal rotation, then $K \cap UK$ with high probability over the choice of U, has diameter bounded by an n-independent constant C.

So the intersection cuts off all outliers and leaves us with the bulk, which is essentially a Euclidean ball.

big theorems (informal versions)

Dvoretzky's thm.: Any normed space $(\mathbb{R}^n, \|\cdot\|)$ has an approximately Euclidean subspace of dim. $\sim \log n$. Moreover, this holds for a randomly chosen subspace of that dimension with prob. $\to 1$ as $n \to \infty$.

[60's & 70's by Dvoretzky & Milman]

(This answered a question by Grothendieck)

Johnson-Lindenstrauss Flattening Lemma: If $\{x_1, \dots, x_n\} =: X \subseteq \ell_2^N$, then there exists an "almost isometric embedding" of X into ℓ_2^k with $k \sim \log n$.

['84 by Johnson & Lindenstrauss]

(This is motivated by data compression and has various applications in comb. optim., learning theory, information retrieval, compressed sensing, data streaming, ...)

II. Normed spaces & convex bodies

Def.: Let X be a vector space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

- A "norm" $\|\cdot\| : X \to \mathbb{R}$ is a function that satisfies $\forall x, y \in X$:

  (i) $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$

  (ii) $\|\lambda x\| = |\lambda| \, \|x\| \quad \forall \lambda \in \mathbb{K}$.

  (iii) $\|x+y\| \leq \|x\| + \|y\|$

  $(X, \|\cdot\|)$ is then a "normed space" and called "Banach space" iff it is complete (i.e., alle Cauchy sequences converge).

Def.:
- Two Banach spaces are "isomorphic" iff there is a linear homeomorphism between them. If this can in addition be chosen to be an isometry, they are called "isometrically isomorphic" or just "isometric" (and we write $X \cong Y$).
- If $T : (X, \|\cdot\|_X) \to (Y, \|Y\|_Y)$ is a linear map between two Banach spaces, then $\|T\| := \sup\limits_{x \in X \setminus \{0\}} \frac{\|T(x)\|_Y}{\|x\|_X} = \sup\limits_{\|x\|_X = 1} \|T(x)\|_Y$. is the "operator norm".
- T is called "bounded" iff $\|T\| < \infty$ and the space of all bounded linear maps will be denoted by $\mathcal{B}(X, Y)$.

Prop.: Let X, Y, Z be Banach spaces and $T \in \mathcal{B}(X,Y)$, $S \in \mathcal{B}(Y,Z)$.

(i) $\|S T\| \leq \|S\| \, \|T\|$

(ii) $\|T(x)\| \leq \|T\| \, \|x\| \quad \forall x \in X$

(iii) An isomorphism T is an isometry iff $\|T\| = \|T^{-1}\| = 1$.

Prop.: In finite dimensions

(i) all normed spaces over $\mathbb{R}$ or $\mathbb{C}$ are Banach spaces,

(ii) all linear maps are bounded (and thus continuous),

(iii) two Banach spaces are isomorphic iff they have the same dimension.

<u>Example:</u> $\ell_p^n := (\mathbb{R}^n, \|\cdot\|_p)$ where

$$\|x\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad \text{for } p \in [1,\infty)$$

$$\|x\|_\infty := \max_i |x_i|$$

Hölder inequality: $\sum_{i=1}^n |x_i y_i| \leq \|x\|_p \|y\|_q$ if $p \in [1,\infty]$ & $\frac{1}{p} + \frac{1}{q} = 1$.

The norm $\|\cdot\|_p$ is induced by a scalar product iff $p=2$. Then Hölder's inequality essentially becomes Cauchy Schwarz.

We will almost exclusively deal with real vector spaces of finite (albeit large) dimension.

<u>Def.:</u> The <u>"dual space"</u> X^* of a normed space X is the space of all cont. linear functionals on X.

<u>Prop.:</u>
- With $X^* \ni f \mapsto \|f\| := \sup_{x \in X \setminus \{0\}} \frac{|f(x)|}{\|x\|}$, X^* becomes a Banach space.
- For $X = (\mathbb{R}^n, \|\cdot\|)$, X^* is isometrically isomorphic to $(\mathbb{R}^n, \|\cdot\|_*)$ with $\|y\|_* = \sup_{\|x\|=1} \langle x, y\rangle$ (and we will in the following identify those spaces).

<u>Example:</u> For $\frac{1}{p} + \frac{1}{q} = 1$ $(\ell_p^n)^* \cong \ell_q^n$.

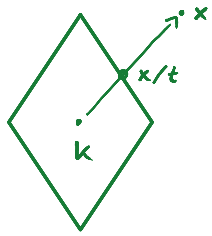
<u>Def.:</u>
- $K \subseteq \mathbb{R}^n$ is <u>"convex"</u> iff $x, y \in K \wedge \lambda \in [0,1] \Rightarrow \lambda x + (1-\lambda) y \in K$
- For $A \subseteq \mathbb{R}^n$, the <u>"convex hull"</u> $\mathrm{conv}(A)$ is the smallest convex set containing A.
- $B \subseteq \mathbb{R}^n$ is a <u>"convex body"</u> iff it is convex, compact and has nonempty interior.
- We call $S \subseteq \mathbb{R}^n$ <u>"symmetric"</u> iff $S = -S$ (i.e. $x \in S \Rightarrow -x \in S$)
- For $C \subseteq \mathbb{R}^n$ the <u>"polar"</u> is $C^\circ := \{ y \in \mathbb{R}^n \mid \langle x, y\rangle \leq 1 \ \forall x \in C\}$.

<u>Examples:</u>
- If C is a convex polytope with v vertices and m faces, then C° is a convex polytope with m vertices and v faces.
- If C is an ellipsoid with semiaxes $r_1, \dots, r_n > 0$, then C° is an ellipsoid with semiaxes $r_1^{-1}, \dots, r_n^{-1}$.

Thm.: (Bipolar thm.) If $S \subseteq \mathbb{R}^n$ is a closed convex set containing the origin, then $S^{\circ\circ} = S$.

Prop.: If $X := (\mathbb{R}^n, \|\cdot\|)$ is a Banach space, then its unit ball $B_X := \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ is a symmetric convex body.
Conversely, if $K \subseteq \mathbb{R}^n$ is a symmetric convex body, then there is a norm $\|\cdot\|_K$ s.t. K is the corresponding unit ball.

proof is elementary and for the converse uses the "Minkowski functional" $\|x\|_K := \inf\{t > 0 \mid x \in tK\}$. $\square$

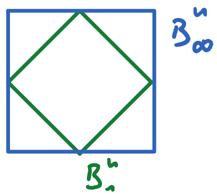


This correspondence is as usefull as simple: it allows to transfer statements from Banach space theory (i.e. functional analysis) to convex geometry & back.

Prop.: Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a Banach space and B_X its unit ball. Then $B_{X^*} = B_X^\circ$.

proof: $B_{X^*} = \{y \in \mathbb{R}^n \mid \underbrace{\|y\|_{X^*}}_{= \sup_{\|x\|=1} \langle x, y\rangle} \leq 1\} = \{y \in \mathbb{R}^n \mid \forall x \in B_X : \langle x, y\rangle \leq 1\} = B_X^\circ$. $\square$

Example: The duality $(\ell_1^n)^* \cong \ell_\infty^n$ is reflected by the fact that $(B_1^n)^\circ = B_\infty^n$



By the bipolar thm. we have that $(B_\infty^n)^\circ = B_1^n$, which in turn corresponds to $(\ell_\infty^n)^* \cong \ell_1^n$.

II. Lewis' thm. & Banach-Mazur distance

Def.: Let $X = Y = \mathbb{R}^n$ and $\|\cdot\|$ be a norm on $\mathcal{B}(X,Y)$. We define the "dual norm" $\|\cdot\|^*$ on $\mathcal{B}(Y,X)$ (by "trace duality") as $\|T\|^* := \sup\{ \operatorname{tr}[ST] \mid \|S\| \le 1\}$.

remark: · By Riesz thm. we have that the dual space $(\mathcal{B}(X,Y), \|\cdot\|)^*$ can be identified (i.e., is isometrically isomorphic) to $(\mathcal{B}(Y,X), \|\cdot\|^*)$ since the space is finite dimensional and $(T_1, T_2) \mapsto \operatorname{tr}[T_1^* T_2]$ a scalar product.

Lemma: Let $X = Y = \mathbb{R}^n$ be Banach spaces and $\|\cdot\|$ the operator norm on $\mathcal{B}(X,Y)$. Then $\|T\|^*$ equals the so-called "nuclear norm"

$$\nu(T) := \min\left\{ \sum_{k=1}^N \|x_k\|_X \|y_k^*\|_{Y^*} \;\middle|\; T = \sum_{k=1}^N x_k y_k^* \text{ where } y_k^* \in Y^*, x_k \in X \text{ and } N \in \mathbb{N} \right\}$$

Moreover, one can choose $N \le n^2$.

remark: Here y_k^* stands for an element in the dual Y^* and, at the same time, for the row vector corresponding to the column vector y_k. Hence,
$y_k^*(y) = \langle y_k, y\rangle$ and
$(x_k y_k^*) : Y \to X$, $y \mapsto x_k y_k^*(y) = x_k \langle y_k, y\rangle$.

proof (sketch): Note that for $S \in \mathcal{B}(X,Y)$:

$$\|S\| = \sup_{\|x\|=1} \|S(x)\| = \sup_{\|x\| = \|y^*\| = 1} y^*(Sx) = \sup\{\operatorname{tr}[Sxy^*] \mid \nu(xy^*) \le 1\}$$
$$= \sup\{\operatorname{tr}[ST] \mid \nu(T) \le 1\}.$$

Hence $B_{(\mathcal{B}(X,Y),\|\cdot\|)} = B^\circ_{(\mathcal{B}(Y,X),\nu)}$ and by the bipolar thm.

$$B_{(\mathcal{B}(Y,X),\nu)} = B^\circ_{(\mathcal{B}(X,Y),\|\cdot\|)} = \{T \in \mathcal{B}(Y,X) \mid \operatorname{tr}[ST] \le 1, \ \forall S \in \mathcal{B}(X,Y) : \|S\| \le 1\}.$$
$$= B_{(\mathcal{B}(X,Y),\|\cdot\|)^*} = B_{(\mathcal{B}(Y,X),\|\cdot\|^*)}.$$

The bound $N \le n^2$ can be shown using Caratheodory's thm. $\square$

Lewis thm.: [Lewis '79] Let $X = (\mathbb{R}^n, \|\cdot\|_X)$ be a Banach space, B_X its unit ball and $\|\cdot\|$ a norm on $\mathcal{B}(X, \mathbb{R}^n)$. If $T \in \mathcal{B}(X, \mathbb{R}^n)$ maximizes $\mathrm{vol}(TB_X)$ among all operators with $\|T\| \le 1$, then T is invertible, $\|T^{-1}\|^* = n$ and $\|T\| = 1$.

remarks:
- $TB_X \subseteq \mathbb{R}^n$ is a convex set and thus always (Lebesgue) measurable.
- Since the $\mathcal{B}(X, \mathbb{R}^n) \simeq \mathbb{R}^{n^2}$ is finite dimensional, its unit ball is compact so that it always contains a T that maximizes $\mathrm{vol}(TB_X)$.

proof: Using that $\mathrm{vol}(TB_X) = |\det(T)|\,\mathrm{vol}(B_X)$ together with the assumed extremal property of T we get $\forall S \in \mathcal{B}(X, \mathbb{R}^n)$:

$$\left|\det\left(\frac{T+S}{\|T+S\|}\right)\right| \le |\det(T)|$$

That implies that T is invertible and we can divide by $|\det(T)|$ to obtain

$$|\det(\mathbb{1} + T^{-1}S)| \le \|T+S\|^n \underset{\uparrow \atop \|T\|\le 1 \;\&\; \text{triangle ineq.}}{\le} (1 + \|S\|)^n$$

Since S is arbitrary we have $\forall \varepsilon > 0$:

$$|\det(\mathbb{1} + \varepsilon T^{-1}S)| \le (1 + \varepsilon\|S\|)^n$$

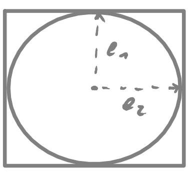
Using that tr is the derivative of $\det$ at $\mathbb{1}$, i.e., $\det(\mathbb{1} + \varepsilon M) = 1 + \varepsilon\,\mathrm{tr}[M] + o(\varepsilon)$ as $\varepsilon \to 0$ we get $\mathrm{tr}[T^{-1}S] \le n\|S\|$ and thus

$$\|T^{-1}\|^* = \sup\{\mathrm{tr}[T^{-1}S] \mid \|S\| \le 1\} \le n.$$

Here equality has to hold since $n = \mathrm{tr}[T^{-1}T] \le \underbrace{\|T^{-1}\|^*}_{\le n}\underbrace{\|T\|}_{\le 1}$. $\square$

Def.: We call $\mathcal{E} \subseteq \mathbb{R}^n$ an <u>"ellipsoid"</u> iff it is the image of $B_{l_2^n}$ under an isomorphism.

example: $X = l_2^n$, $Y = l_\infty^n$. Then B_X is the largest ellipsoid in B_Y and for the standard ONB $\{e_1, \dots, e_n\}$ we have that $\forall k$: $e_k \in \partial B_X \cap \partial B_Y$ and $\mathbb{1} = \sum_{k=1}^n e_k e_k^*$.



This is generalized in the following:

Cor.: Let $X=(\mathbb{R}^n,\|\cdot\|_X)$, $Y=(\mathbb{R}^n,\|\cdot\|_Y)$ with unit balls s.t. $B_X \subseteq B_Y$ and $\mathrm{vol}(TB_X)\le \mathrm{vol}(B_X)$ for every $T\in\mathcal{B}(X,Y)$ for which $TB_X\subseteq B_Y$.
Then there are points $\{x_1,\dots,x_N\}\subseteq \partial B_X\cap\partial B_Y$, $\{y_1^*,\dots,y_N^*\}\subseteq \partial B_{X^*}\cap\partial B_{Y^*}$ and $\lambda_k>0$ s.t. $\mathbb{1}=\sum_{k=1}^N \lambda_k\, x_k y_k^*$ and $\sum_{k=1}^N \lambda_k = n$.
Moreover, one can choose $N\le n^2$.

proof: We apply Lewis' theorem with the operator norm $\|\cdot\|$ on $\mathcal{B}(X,Y)$. Then $TB_X\subseteq B_Y$ corresponds to $\|T\|\le 1$ and the identity map by assumption maximizes $\mathrm{vol}(TB_X)$ among all maps with $\|T\|\le 1$. Hence

$$n=\|\mathbb{1}\|^*=\min\Big\{\sum_{k=1}^N \|a_k\|_X\|b_k^*\|_{Y^*}\ \Big|\ \mathbb{1}=\sum_{k=1}^N a_k b_k^*\Big\}$$
$$=\min\Big\{\underbrace{\sum_{k=1}^N \lambda_k\|x_k\|_X\|y_k^*\|_{Y^*}}_{=\sum_{k=1}^N\lambda_k}\ \Big|\ \mathbb{1}=\sum_{k=1}^N\lambda_k x_k y_k^* \wedge \{x_k\}\in\partial B_X \wedge \{y_k^*\}\in\partial B_{Y^*}\wedge \lambda\in\mathbb{R}_+^N\Big\}$$

where we used that the nuclear norm is dual to the operator norm.

W.l.o.g. we can choose $\lambda_k>0$ and $N\le n^2$.

Since $B_X\subseteq B_Y$ and thus $\|x_k\|_Y\le\|x_k\|_X$ we have

$$n=\mathrm{tr}[\mathbb{1}]=\sum_{k=1}^N\lambda_k\langle x_k,y_k\rangle\le\sum_{k=1}^N\lambda_k|\langle x_k,y_k\rangle|$$
$$\le\sum_{k=1}^N\lambda_k\|x_k\|_Y\|y_k^*\|_{Y^*}\le\sum_{k=1}^N\lambda_k\|x_k\|_X\|y_k^*\|_{Y^*}=\sum_{k=1}^N\lambda_k=n.$$

So $\forall k$: $\|x_k\|_X=\|x_k\|_Y=1$ and $\langle x_k,y_k\rangle=1$.

From the latter we get $1=\langle x_k,y_k\rangle\le\underbrace{\|x_k\|_X}_{=1}\|y_k^*\|_{X^*}\underset{\substack{\uparrow\\ B_{Y^*}\subseteq B_{X^*}\\ \text{exercise}}}{\le}\|y_k^*\|_{Y^*}=1$,
which implies $\|y_k\|_{X^*}=1$. $\square$

remarks:
- The x_k's and y_k^*'s are often called "contact points" since the unit balls touch there.
- In the exercise we will see that this corollary has a simple converse.
- If X or Y is ℓ_2^n in the previous corollary, then $y_k=x_k$ since then $\langle x_k,y_k\rangle=\|x_k\|_{\ell_2^n}\|y_k\|_{\ell_2^n}$.

John's thm. [John '48]:

Let $Y = \mathbb{R}^n$ be a Banach space and B_Y its unit ball. There is a unique maximum volume ellipsoid $\mathcal{E}$ inside B_Y and it holds that
$\mathcal{E} \subseteq B_Y \subseteq \sqrt{n}\, \mathcal{E}$.

proof: We first prove uniqueness.
Let $\mathcal{E}$ be the image of $B_{\ell_2^n}$ under some $T \in \mathcal{B}(\ell_2^n, Y)$. Then $\mathcal{E} \subseteq B_Y$ is equivalent to $\|T\| \leq 1$ for the operator norm. Suppose the volume is maximized for T_1 and T_2. By Lewis' thm. $\|T_i\| = 1$ and $\|T_i^{-1}\|^* = n$. By polar decomposition there is an orthogonal V s.t. $P := T_2^{-1} T_1 V \geq 0$. Let $\{p_i > 0\}_{i=1}^n$ be the eigenvalues of P. Then
$\sum_{i=1}^n p_i = \mathrm{tr}[P] \leq \|T_2^{-1}\|^* \|T_1\| = n$ and $\sum_{i=1}^n \frac{1}{p_i} = \mathrm{tr}[P^{-1}] \leq \|T_1^{-1}\|^* \|T_2\| = n$.
Hence $p_i = 1 \ \forall i$ (since $n = \langle p^{1/2}, p^{-1/2} \rangle \leq \|p^{1/2}\| \, \|p^{-1/2}\| = n$)
so that $T_2 = T_1 V$ which implies uniqueness of $\mathcal{E}$ since $V B_{\ell_2^n} = B_{\ell_2^n}$.
For the second part, assume w.l.o.g. that $\mathcal{E} = B_{\ell_2^n}$ (by applying a suitable isomorphism). Then take any $z \in B_Y$ and use the decomposition $\mathbb{1} = \sum_k \lambda_k x_k x_k^*$ from the previous corollary:

$$\|z\|^2 = \langle z, z \rangle = \sum_k \lambda_k |\langle z, x_k \rangle|^2 \leq \sum_k \lambda_k \underbrace{\|z\|_Y^2}_{\leq 1} \underbrace{\|x_k\|_{Y^*}^2}_{= 1} \leq n .$$

So $B_Y \subseteq \sqrt{n}\, \mathcal{E}$. $\square$

That the $\sqrt{n}$ factor in John's thm. is sharp, is seen be taking $Y = \ell_\infty^n$. Then (as proven in the exercise) $\mathcal{E} = B_{\ell_2^n}$ is the max. volume ellipsoid and since $e := (1, \dots, 1)$ satisfies $e \in B_Y$ and $\|e\|_{\ell_2^n} = \sqrt{n}$ we have that $B_Y \nsubseteq c\, \mathcal{E}$ for any $c < \sqrt{n}$.

<u>Def.:</u> • Let X, Y be two isomorphic normed spaces. The (analytic) "Banach-Mazur distance" is defined as

$$d(X,Y) := \inf \{ \|T\| \|T^{-1}\| \mid T: X \to Y \text{ isomorphism} \}$$

• Let K, L be symmetric convex bodies in $\mathbb{R}^n$. The (geometric) "Banach-Mazur distance" is defined as

$$d(K,L) := \inf \{ ab > 0 \mid \tfrac{1}{b} L \subseteq TK \subseteq aL \text{ for some } T \in B(\mathbb{R}^n) \}$$

<u>Cor.:</u> If $X \simeq Y \simeq \mathbb{R}^n$ are normed spaces with unit balls B_X and B_Y and $R: X \to Y$ is an isomorphism, then $d(X,Y) = d(B_X, B_Y) = d(RB_X, RB_Y)$.

<u>proof:</u> The first equality follows from $\|T\| = \inf \{ a \in \mathbb{R} \mid \forall x \in X: \frac{\|Tx\|}{\|x\|} \leq a \}$

$$= \inf \{ a \in \mathbb{R} \mid TB_X \subseteq aB_Y \}$$

and the second from $TB_X \subseteq aB_Y \Leftrightarrow T'RB_X \subseteq aRB_Y$ with $T' := RTR^{-1}$. □

<u>Prop.:</u> (Properties of Banach-Mazur distance)

Let $X \simeq Y \simeq Z \simeq \mathbb{R}^n$ be normed spaces. Then

(i) $d(X,Y) \geq 1$ and $d(X,Y) = 1$ iff $X \cong Y$.

(ii) $d(X,Y) = d(Y,X)$

(iii) $d(X,Y) \leq d(X,Z)\, d(Z,Y)$

(iv) $d(X,Y) = d(X^*, Y^*)$.

- These follow easily from the definitions and the properties of the operator norm.
- Note that after a logarithm, (i), (ii) and (iii) become non-negativity, symmetry and triangle inequality of a metric:

<u>Cor.:</u> (Banach Mazur metric) $(X,Y) \mapsto \log d(X,Y)$ defines a metric on the space of all Banach spaces that are isomorphic to $\mathbb{R}^n$ when isometric spaces are identified.

<u>remarks:</u> • Equivalently, one obtains a metric on the quotient space of symmetric convex bodies in $\mathbb{R}^n$ under the equivalence relation $K \sim L \Leftrightarrow \exists M \in GL(n): K = ML$.

That is, ellipsoid and balls are identified as well as cubes and parallelotopes, but not cubes and balls.

- The obtained metric space turns out to be compact and is called the "Banach-Mazur compactum".
- $d(X,Y)$ is usually difficult to compute.

Thm: (John's thm. 2^{nd} version)

Let $X \simeq Y \simeq \mathbb{R}^n$ be normed spaces. Then $d(X, \ell_2^n) \le \sqrt{n}$ and $d(X,Y) \le n$.

proof: From the previously stated version of John's thm. we know that there is an isomorphism $T: \ell_2^n \to X$ s.t. with $\mathcal{E} := T B_{\ell_2^n}$ we have $\mathcal{E} \subseteq B_X \subseteq \sqrt{n}\, \mathcal{E}$. Hence $B_{\ell_2^n} \subseteq T^{-1} B_X \subseteq \sqrt{n}\, B_{\ell_2^n}$ and thus $d(X, \ell_2^n) \le \sqrt{n}$.

This implies the second statement since $d(X,Y) \le d(X, \ell_2^n)\, d(\ell_2^n, Y) \le n$.

$\square$

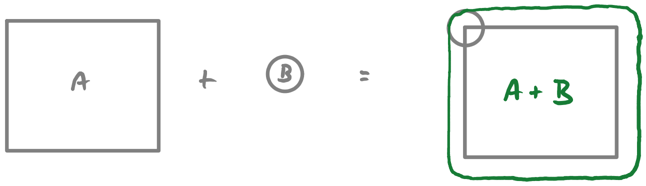
remarks:
- From the discussion about sharpness of John's thm. we get $d(\ell_\infty^n, \ell_2^n) = \sqrt{n}$.
- More generally, if $1 \le p \le q \le 2$ or $2 \le p \le q \le \infty$, then $d(\ell_p^n, \ell_q^n) = n^{1/p - 1/q}$ ($\to$ exercise).

  In this case $T = id$ turns out to be optimal. For $1 \le p \le 2 < q \le \infty$, however, the situation is more complicated.
- In 1981 Gluskin came up with a non-trivial construction of n-dim. spaces X and Y s.t. $d(X,Y) \ge cn$ proving that $d(X,Y) \le n$ is sharp up to a constant.

III. Brunn-Minkowski inequality

Def.: If $A, B \subseteq \mathbb{R}^n$, then the "Minkowski sum" is defined as
$A + B := \{ a+b \mid a \in A \wedge b \in B \}$.

Example: $A = B_{\ell_\infty^n}$, $B := \varepsilon B_{\ell_2^n}$

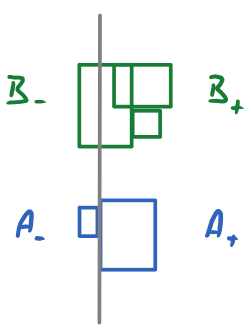


Thm.: (Brunn-Minkowski)
Let $A, B \subseteq \mathbb{R}^n$ be compact with nonzero volume and $\lambda \in [0,1]$. Then
(i) $\mathrm{vol}(A+B)^{1/n} \geq \mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}$ and
(ii) $\mathrm{vol}(\lambda A + (1-\lambda) B) \geq \mathrm{vol}(A)^{\lambda} \mathrm{vol}(B)^{1-\lambda}$. ($\sim$ AGM for the volume)

proof: (i) If A & B are boxes with edge lengths $a, b \in \mathbb{R}^n_{>0}$, then $A+B$ is also a box, now with edge lengths $a+b$. In this case (i) is equivalent to $\left(\prod_{i=1}^n \frac{a_i}{a_i+b_i} \right)^{1/n} + \left(\prod_{i=1}^n \frac{b_i}{a_i+b_i} \right)^{1/n} \leq 1$. This follows from the arithmetic-geometric mean inequality, which implies that the l.h.s. is upper bounded by $\frac{1}{n} \sum_{i=1}^n \frac{a_i}{a_i+b_i} + \frac{1}{n} \sum_{i=1}^n \frac{b_i}{a_i+b_i} = 1$.

Next, we will prove (i) for a collection of boxes by induction. Let $\#(A \cup B)$ be the total number of boxes in A and B & assume that (i) holds for $\#(A \cup B) = m-1$ (with $\#(A \cup B) = 2$ just proven). We now prove it for $\#(A \cup B) = m$.
Now translate A & B separately, s.t. $A = A_+ \cup A_-$ and $B = B_+ \cup B_-$ where $A_+, B_+ / A_-, B_-$ are the intersections of A, B with the positive/negative half space and $\max\{ \#(A_+ \cup B_+), \#(A_- \cup B_-) \} \leq m-1$ and

$$\frac{\mathrm{vol}(A_\pm)}{\mathrm{vol}(A)} = \frac{\mathrm{vol}(B_\pm)}{\mathrm{vol}(B)}.$$



Since $A+B = (A_+ + B_+) \cup (A_+ + B_-) \cup (A_- + B_+) \cup (A_- + B_-)$

we have

$$\begin{aligned}
\mathrm{vol}(A+B) &\geq \mathrm{vol}(A_+ + B_+) + \mathrm{vol}(A_- + B_-) \\
&\geq \left(\mathrm{vol}(A_+)^{1/n} + \mathrm{vol}(B_+)^{1/n}\right)^n + \left(\mathrm{vol}(A_-)^{1/n} + \mathrm{vol}(B_-)^{1/n}\right)^n \quad \text{by induction hypothesis} \\
&= \mathrm{vol}(A_+)\left(1 + \frac{\mathrm{vol}(B)^{1/n}}{\mathrm{vol}(A)^{1/n}}\right)^n + \mathrm{vol}(A_-)\left(1 + \frac{\mathrm{vol}(B)^{1/n}}{\mathrm{vol}(A)^{1/n}}\right)^n \\
&= \mathrm{vol}(A)\left(1 + \frac{\mathrm{vol}(B)^{1/n}}{\mathrm{vol}(A)^{1/n}}\right)^n = \left(\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}\right)^n.
\end{aligned}$$

This proves the result for finite union of boxes, which can finally be used to approximate arbitrary compact sets A and B.

(ii) follows from (i) by applying it to sets λA and $(1-\lambda)B$:

$$\mathrm{vol}(\lambda A + (1-\lambda)B) \overset{(i)}{\geq} \left(\lambda\, \mathrm{vol}(A)^{1/n} + (1-\lambda)\,\mathrm{vol}(B)^{1/n}\right)^n \underset{\text{arithmetic-geometric mean}}{\geq} \mathrm{vol}(A)^{\lambda}\,\mathrm{vol}(B)^{1-\lambda}$$

$\square$

<u>remark</u>: • One can also derive (i) from (ii) by setting $\lambda := \dfrac{\mathrm{vol}(A)^{1/n}}{\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}}$.

Then with $A' := A/\mathrm{vol}(A)^{1/n}$, $B' := B/\mathrm{vol}(B)^{1/n}$ (ii) implies:

$\mathrm{vol}(\lambda A' + (1-\lambda)B') \geq \mathrm{vol}(A')^{\lambda}\,\mathrm{vol}(B')^{1-\lambda} = 1$, which in turn implies (i)

since $\lambda A' + (1-\lambda)B' = (A+B)/(\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n})$.

<u>Cor.:</u> <u>(Isoperimetric inequality)</u>

If $A \subseteq \mathbb{R}^n$ is a convex body and $B := B_{l_2^n}$. Then $\left(\dfrac{\mathrm{vol}(A)}{\mathrm{vol}(B)}\right)^{1/n} \leq \left(\dfrac{S(A)}{S(B)}\right)^{\frac{1}{n-1}}$

when $S(K) := \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon}\left(\mathrm{vol}(K+\varepsilon B) - \mathrm{vol}(K)\right)$ defines the surface area.

(In particular, if $S(A) = S(B)$, then $\mathrm{vol}(A) \leq \mathrm{vol}(B)$)

<u>proof:</u>

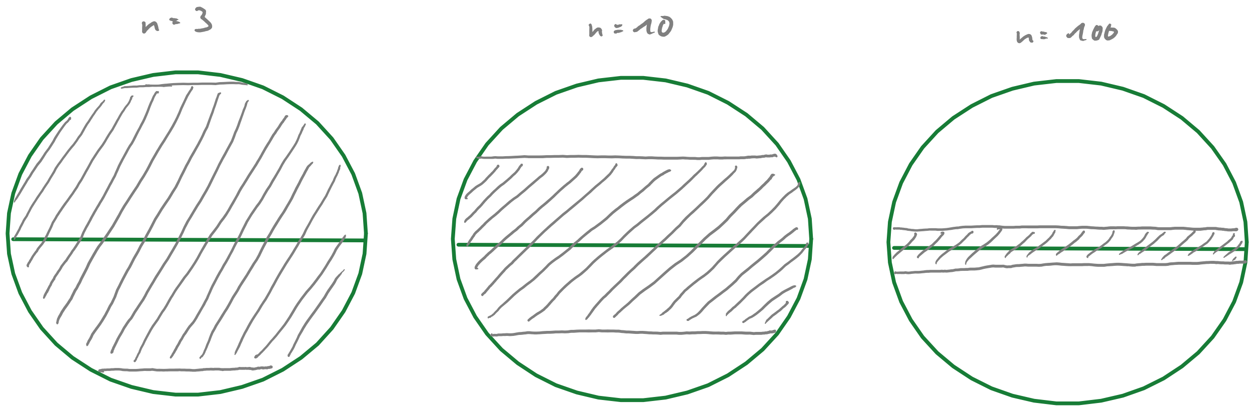
$$\begin{aligned}
S(A) &\overset{\text{B.M.}}{\geq} \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon}\left[\left(\mathrm{vol}(A)^{1/n} + \varepsilon\,\mathrm{vol}(B)^{1/n}\right)^n - \mathrm{vol}(A)\right] \\
&= n\,\mathrm{vol}(A)^{\frac{n-1}{n}}\,\mathrm{vol}(B)^{\frac{1}{n}} \underset{S(B) = n\,\mathrm{vol}(B)}{=} \mathrm{vol}(A)^{\frac{n-1}{n}}\,\mathrm{vol}(B)^{\frac{1-n}{n}}\,S(B).
\end{aligned}$$

$\square$

IV. Concentration of measure on S^n

Def.: • We denote by $S^{n-1} := \{ x \in \mathbb{R}^n \mid \|x\|_2 = 1 \}$ the unit sphere in $\mathbb{R}^n$, which we regard as a metric space with metric $(x,y) \mapsto \|x-y\|_2$.
(We could as well take the geodesic distance $d(x,y) := \arccos \langle x,y \rangle$, which satisfies $\frac{\pi}{2}\|x-y\|_2 \geq d(x,y) \geq \|x-y\|_2$)

• By σ we denote the normalized "spherical measure" on S^{n-1}.
That is, σ is defined on Borel subsets of S^{n-1} and satisfies $\sigma(S^{n-1}) = 1$ and $\sigma(TA) = \sigma(A)$ $\forall T \in O(n)$.
(This measure exists, is unique and satisfies
$\sigma(A) = vol(\tilde{A}) / vol(B_{\ell_2^n})$ with $\tilde{A} := \{ x \in \mathbb{R}^n \mid \exists \lambda \in [0,1] \, \exists a \in A : x = \lambda a \}$)



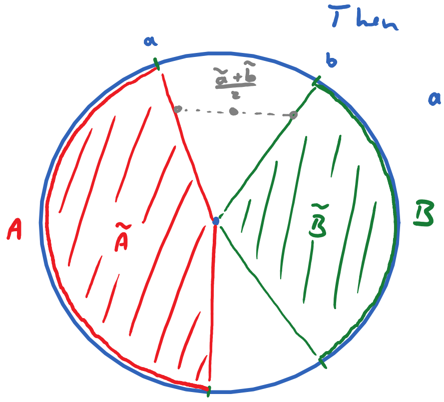
A band around the "equator" that covers 90% of the sphere shrinks as $O(1/\sqrt{n})$ as the dimension n is increased. That is, the surface measure "concentrates" around the equator.

We will see a geometric & a functional version of concentration of measure on the sphere. Here's the geometric one:

Thm.: Let $A \subseteq S^{n-1}$ be a measurable subset and $A_\delta := \{ x \in S^{n-1} \mid \exists y \in A : \|x-y\|_2 \leq \delta \}$. Then

$$\sigma(A_\delta) \geq 1 - \frac{e^{-n\delta^2/4}}{\sigma(A)} .$$

<u>proof:</u> Define $B := S^{n-1} \setminus A_\delta$, $\tilde{A} := [0,1]\cdot A$, $\tilde{B} := [0,1]\cdot B$ as on p. 14.



Then $vol(\tilde{A}) = \sigma(A_\delta) \cdot vol(B_{\ell_2^n})$, $vol(\tilde{B}) = \sigma(B)\, vol(B_{\ell_2^n})$

and for any $\tilde{a} \in \tilde{A}$, $\tilde{b} \in \tilde{B}$ with $a := \frac{\tilde{a}}{\|\tilde{a}\|_2}$, $b := \frac{\tilde{b}}{\|\tilde{b}\|_2}$ we can bound $\left\| \frac{\tilde{a}+\tilde{b}}{2} \right\| \le \left\| \frac{a+b}{2} \right\| \le \sqrt{1-\frac{\delta^2}{4}} =: r$.

Hence, $\frac{1}{2}(\tilde{A}+\tilde{B}) \subseteq r B_{\ell_2^n}$ so that

$$vol(\tilde{A})^{1/2} vol(\tilde{B})^{1/2} \overset{B.M.}{\le} vol\left(\frac{\tilde{A}+\tilde{B}}{2}\right) \le r^n vol(B_{\ell_2^n})$$

$$= \sigma(A)^{1/2} \sigma(B)^{1/2} vol(B_{\ell_2^n}),$$

which implies

$$1-\sigma(A_\delta) = \sigma(B) \le \sigma(A)^{-1} r^{2n}$$

$$= \sigma(A)^{-1}\left(1-\frac{\delta^2}{4}\right)^n \underset{1+x \le e^x}{\le} \sigma(A)^{-1} e^{-n\delta^2/4}. \quad \square$$

Taking A as the upper hemisphere, and thus $\sigma(A) = \frac{1}{2}$, we get that e.g. 99% of the sphere is within a distance of $O(\frac{1}{\sqrt{n}})$ from A. Repeating the argument with the lower hemisphere and combining the two, we get that 99% are within distance $O(\frac{1}{\sqrt{n}})$ from the equator.

Concentration of measure also occurs in other spaces. One example:

<u>Gaussian measure concentration:</u>

For $A \subseteq \mathbb{R}^n$ let $\gamma(A) := (2\pi)^{-n/2} \int_A e^{-\|x\|^2/2} dx$ be the Gaussian measure. Then for any measurable A with $\gamma(A) \ge \frac{1}{2}$:

$$\gamma(A_\delta) \ge 1 - e^{-\delta^2/2}.$$

Here n does not enter the bound directly, but note that a random point x (according to γ) has $\|x\| \sim \sqrt{n}$.

Other concentration of measure results are known for the n-dim torus $S^1 \times \dots \times S^1$, the Hamming cube $\{0,1\}^n$, etc. Most of them can be proven from an isoperimetric inequality.

Since σ is a probability measure on S^{n-1}, any measurable function $f: S^{n-1} \to \mathbb{R}$ can be regarded as random variable with

"expectation value" $\quad E(f) := \int_{S^{n-1}} f(x)\, d\sigma(x)$ and a

"median" $\quad \mathrm{med}(f) := \sup\{ t \in \mathbb{R} \mid P[f \le t] \le \frac{1}{2} \}$ where

$$P[f \le t] := \sigma(\{ x \in S^{n-1} \mid f(x) \le t \}).$$

Note that by construction $P[f < \mathrm{med}(f)] \le \frac{1}{2}$ and $P[f > \mathrm{med}(f)] \le \frac{1}{2}$.

The following is a 'functional version' of concentration of measure on the sphere:

Thm.: (Lévy's Lemma) If $f: S^{n-1} \to \mathbb{R}$ is a function that satisfies $|f(x) - f(y)| \le \|x - y\|$ (i.e., it is "1-Lipschitz"), then $\forall \delta \in [0,1]$:

$$P[f > \mathrm{med}(f) + \delta] \le 2e^{-\delta^2 n/4} \quad \text{and}$$

$$P[f < \mathrm{med}(f) - \delta] \le 2e^{-\delta^2 n/4}.$$

proof: Define $A := \{ x \in S^{n-1} \mid f(x) \le \mathrm{med}(f) \}$. Then $\sigma(A) \ge \frac{1}{2}$ and if $y \in A_\delta$, then $\exists x \in A$ with $\delta \ge \|y - x\|_2 \ge |f(y) - f(x)|$ so that $y \in A_\delta \Rightarrow f(y) \le \mathrm{med}(f) + \delta$.

Using the previous thm. we get

$$P[f > \mathrm{med}(f) + \delta] \le 1 - \sigma(A_\delta) \le 2e^{-n\delta^2/4}.$$

The second inequality is proven in the same way by choosing A accordingly. $\square$

Here we could as well have taken the expectation value instead of the median:

Lemma: Let $f: S^{n-1} \to \mathbb{R}$ be a measurable 1-Lipschitz function. Then
$$|\,\mathrm{med}(f) - \mathbb{E}(f)\,| \leq 4\frac{\sqrt{\pi}}{\sqrt{n}} .$$

proof: $m := \mathrm{med}(f)$

$$|\,m - \mathbb{E}(f)\,| = |\,\mathbb{E}(m-f)\,| \leq \mathbb{E}(\,|m-f|\,) \qquad , \; g(x) := |m - f(x)|$$
$$= \int_{S^{n-1}} g(x)\, d\sigma(x)$$
$$= \int_{S^{n-1}} \left(\int_0^\infty \chi(x,t)\, dt \right) d\sigma(x) \qquad , \; \chi(x,t) := \begin{cases} 1, & t < g(x) \\ 0 \end{cases}$$
$$\overset{\text{Fubini}}{=} \int_0^\infty \left(\int_{S^{n-1}} \chi(x,t)\, d\sigma(x) \right) dt$$
$$= \int_0^\infty P[\,|m-f| > t\,]\, dt$$
$$\leq \int_0^\infty 4 e^{-t^2 n/4}\, dt = \frac{4\sqrt{\pi}}{\sqrt{n}} . \qquad \square$$

V. Johnson-Lindenstrauss Lemma

Thm.: (Existence of Haar measure)
Let (M,d) be a compact metric space and G a group of isometries $g: M \to M$ (i.e., $\forall x,y \in M\ \forall g \in G:\ d(gx,gy) = d(x,y)$).
Then there exists a probability measure μ on the Borel sets that is G-invariant (i.e., $\forall g \in G\ \forall$ Borel sets $A \subseteq M:\ \mu(gA) = \mu(A)$).
Moreover, μ is unique, if G acts transitively on M (i.e., $\forall x \in M:\ Gx = M$).
This measure is called "Haar measure".

examples:
- $M = S^{n-1}$, $G = O(n)$: Then $\mu = \sigma$ is the spherical measure.
- $M = G = O(n)$: If δ is a metric on S^{n-1} w.r.t. which G is a group of isometries, then $d(g,h) := \sup_{x \in S^{n-1}} \delta(gx, hx)$ defines a metric on G, which inherits this property.
In this case the Haar measure defines "random rotations".
- $G = O(n)$, $M = G_{n,k} := \{ k\text{-dim. subspaces of } \mathbb{R}^n \}$ "Grassmannian"
Each element in $G_{n,k}$ can be identified with a k-dim orth. projection $P \in \mathbb{R}^{n \times n}$. Group action and metric can then be defined as $P \mapsto gPg^{-1}$ and $d(P,P') = \|P-P'\|_{op}$. The Haar measure then defines "random subspaces" or "random projections".

Lemma: Let $U \in O(n)$ be a Haar-random rotation.

(i) Let $x \in S^{n-1}$ be fixed. Then Ux is distributed according to the spherical measure on S^{n-1}.

(ii) Let $L \in G_{n,k}$. Then $U(L)$ is a Haar-random k-dim. subspace in $\mathbb{R}^n$.

proof: (i) Let $A \subseteq S^{n-1}$ be a Borel set, then the probability for $Ux \in A$ satisfies

$$P[Ux \in A] = \mu(\{U \in O(n) \mid Ux \in A\})$$
$$= \mu(\{U \in O(n) \mid V^{-1}Ux \in A\})$$
$$= P[Ux \in VA]$$

inv. of Haar measure μ, $V \in O(n)$

Hence, $P[Ux \in A]$ defines an $O(n)$-invariant measure on Borel sets of S^{n-1}. By uniqueness of the spherical measure we have $P[Ux \in A] = \sigma(A)$.

(ii) $\rightarrow$ exercise. $\square$

Thm.: (Johnson-Lindenstrauss ['84])

For every $\varepsilon > 0$ there is a constant $c > 0$ s.t. for every Hilbert space $\mathcal{H}$ and any subset $X \subseteq \mathcal{H}$ with $|X| =: n \in \mathbb{N}$ there is a linear map $L: \mathcal{H} \to l_2^k$ with $k \le c \log n$ s.t. $\forall u, v \in X$: $(1-\varepsilon)\|u-v\| \le \|Lu - Lv\| \le (1+\varepsilon)\|u-v\|$.

remarks:
- This is referred to as the "Johnson-Lindenstrauss flattening Lemma".
- The result has > 1500 citations.
- The possibility of "dimensionality reduction" of this type depends on the norm. For the $\|\cdot\|_\infty$-norm for instance it is not possible & one can show [Johnson, Naor 2008] that the JL-Lemma almost characterizes Hilbert spaces.
- $\dim(\mathcal{H})$ does not appear in the thm. Note that w.l.o.g. we can assume $\dim(\mathcal{H}) = n$. With this assumption the following Prop. provides more details and implies the JL-Lemma:

<u>Prop.:</u> Let $n > 3$, $\varepsilon \in [0,1]$, $\{u_1, \dots, u_n\} \subseteq \mathbb{R}^n$. If $Q: \mathbb{R}^n \to \mathbb{R}^n$ is a random orth. projection onto a subspace of dimension $k = \frac{64}{\varepsilon^2} \log n$, then with probability $\geq 1 - \frac{1}{n}$ we have $\forall i,j \in \{1, \dots, n\}$:

(*) $$(1-\varepsilon)\,\|u_i - u_j\|_2 \leq \left\| \tfrac{1}{m} Q (u_i - u_j) \right\|_2 \leq (1+\varepsilon)\,\|u_i - u_j\|_2$$

where $m := \operatorname{med} \|Qx\|_2$ for any $x \in S^{n-1}$.

<u>remark:</u> The constants are not optimal. With some effort one can replace 64 by 16. The scaling can be improved to $\sim \frac{\log(n)}{\varepsilon^2 \log(1/\varepsilon)}$.

<u>proof:</u> Our aim is to choose k such that for any $x \in S^{n-1}$ and Haar-random k-dim. orth. projection Q:

(1) $$P\left[\, \big| \|Qx\| - m \big| \geq \varepsilon m \right] < n^{-3}$$

If this is true, then for any pair $u_j \neq u_i$ the equation (*) is violated with probability $< n^{-3}$ and since there are $\binom{n}{2} \leq n^2$ such pairs it holds for all of them with probability $\geq 1 - \frac{1}{n}$.

By the previous Lemma we can instead of fixing $x \in S^{n-1}$ and choosing a random projection in (1) as well fix Q and choose x randomly according to the spherical measure. Then $f: x \mapsto \|Qx\|$ is 1-Lipschitz with median m and Lévy's Lemma gives:

(2) $$P\left[\, \big| \|Qx\| - m \big| \geq \varepsilon m \right] \leq 4 e^{-\varepsilon^2 m^2 n/4}.$$

To proceed, we compute a lower bound for m: to this end, note that

- $$\mathbb{E}(f^2) = \mathbb{E}(\|Qx\|^2) \quad \text{for fixed } x \in S^{n-1} \text{ \& } Q \text{ random}$$
$$= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} \langle x_i, Q x_i \rangle \right) = \frac{1}{n} \operatorname{tr}[Q] = \frac{k}{n}$$

- $$\mathbb{E}(f^2) \leq P[f \leq m+\delta]\,(m+\delta)^2 + P[f > m+\delta] \max_x f(x)^2$$
$$\leq (m+\delta)^2 + 2e^{-\delta^2 n/4} \leq 2(m^2 + \delta^2 + e^{-\delta^2 n/4})$$
$$\underset{\delta = \sqrt{\frac{k}{cn}}}{=} 2\left(m^2 + \frac{k}{cn} + e^{-\frac{k}{4c}}\right) \underset{k > 4c\log n}{\leq} 2\left(m^2 + \frac{k}{cn} + \frac{1}{n}\right)$$

- combining the two: $m^2 \geq \frac{k}{2n} - \frac{k}{cn} - \frac{1}{n} \overset{c=6}{=} \frac{1}{3}\frac{k-3}{n} \overset{k\geq 12}{\geq} \frac{1}{4}\frac{k}{n}$.

Inserting this into the r.h.s. of (2) we get
$P[\,|\,\|Qx\| - m\,| > \varepsilon m\,] \leq 4\, e^{-k\varepsilon^2/16}$,
which is smaller than n^{-3} if $k \geq \frac{16}{\varepsilon^2}\log(4n^3)$.
So for $n \geq 3$ we can choose $k = \frac{64}{\varepsilon^2}\log n$. $\square$

remarks:

- An advantage compared to other dimensionality reduction techniques is that JL is non-adaptive, i.e., the projection can be chosen before the data is available. This allows e.g. to use JL in 'streaming algorithms'.

- Instead of using random projections one can use random constructions based on Gaussian or Bernoulli random variables (used for the entries of Q).

- For each data point $x \in \mathbb{R}^N$ one has to compute Qx which, since Q has $O(\frac{N}{\varepsilon^2}\log n)$ entries, requires $O(\frac{N}{\varepsilon^2}\log n)$ operations. In 2006 Ailon & Chazelle introduced the "Fast JL transform" that has computational cost closer to $O(N \log N)$ and uses Q of the form $Q = PFD$, where P is sparse, F a Fourier trafo and D diagonal with random signs.

- Recall the meaning of the symbols $O(\cdot)$ and $\Omega(\cdot)$ for the asymptotic behavior in some limit, e.g. $n \to \infty$:
  $f(n) = O(g(n))$ means g is up to a constant an asymptotic upper bound for f, i.e. $\limsup_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| < \infty$.
  Equivalently, $\exists C > 0\ \exists m\ \forall n > m: \ |f(n)| \leq C\,|g(n)|$.
  $f(n) = \Omega(g(n))$ means g is up to a constant an asympt. lower bound for f, i.e. $\liminf_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| > 0$.

VI. Metric embedding theory

Def.: Let (X,d), (Y,δ) be metric spaces. An injective map $f:X\to Y$ is called a "D-embedding", where $D\geq 1$ is a real number, if there is an $r>0$ s.t.
$\forall x,y\in X$: $\quad r\,d(x,y)\leq \delta(f(x),f(y))\leq D\,r\,d(x,y)$.
The inf. over all D's for which f is a D-embedding is called "distortion" of f.

remarks:

- With $Lip(f) := \sup_{x\neq y} \frac{\delta(f(x),f(y))}{d(x,y)}$, $Lip(f^{-1}) := \sup_{x\neq y}\frac{d(x,y)}{\delta(f(x),f(y))}$ we get that the distortion is $D = Lip(f)\cdot Lip(f^{-1})$.
- If $X\simeq Y\simeq \mathbb{R}^n$ are normed spaces of Banach-Mazur distance $d(X,Y)$, then there exists a linear D-embedding iff $D\geq d(X,Y)$.
- In JL we have a D-embedding with $D = \frac{1+\varepsilon}{1-\varepsilon} = 1+2\varepsilon+O(\varepsilon^2)$ into l_2^k with $k=O\left(\frac{\log|X|}{\varepsilon^2}\right)$. The following shows that a distortion independent of $|X|$ cannot always be achieved when embedding into l_2:

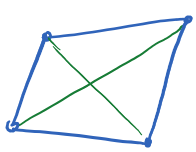
Thm.: [Enflo '69] Let $X:=\{0,1\}^m$, $m\geq 2$ be the "Hamming cube" equipped with the l_1-metric ("Hamming distance"). If $f:X\to l_2$ is a D-embedding, then $D\geq\sqrt{m}$.

proof (sketch): The idea is to show that if one embeds such that edges have approximately the right length, then some diagonals get too short.

$E := \{(u,v)\in X\times X \mid \|u-v\|_1 = 1\}$ 'edges'

$F := \{(u,v)\in X\times X \mid \|u-v\|_1 = m\}$ 'long diagonals'

Note that $|E| = m2^{m-1}$ and $|F| = 2^{m-1}$.



For real numbers $x_1,\dots,x_4$ one has

$(x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_4)^2+(x_4-x_1)^2-(x_1-x_3)^2-(x_2-x_4)^2$

$=(x_1-x_2+x_3-x_4)^2\geq 0$. Hence for $m=2$:

$$\sum_{(u,v)\in F}\|f(u)-f(v)\|_2^2 \leq \sum_{(u,v)\in E}\|f(u)-f(v)\|_2^2 .$$

For $m>2$ the same inequality can be shown by induction (which we omit).

So $$1 \geq \frac{\sum_{(u,v)\in F} \|f(u)-f(v)\|_2^2}{\sum_{(u,v)\in E} \|f(u)-f(v)\|_2^2} \underset{\substack{\uparrow \\ f \text{ is } D\text{-embedding}}}{\geq} \frac{1}{D^2} \frac{\sum_{(u,v)\in F} \|u-v\|_1^2}{\sum_{(u,v)\in E} \|u-v\|_1^2} = \frac{1}{D^2} \frac{m^2 2^{m-1}}{m\, 2^{m-1}} = \frac{m}{D^2}.$$ $\square$

remarks:
- In terms of $|X| =: n$ we obtain that there is no embedding with distortion less than $\sqrt{\log n}$ into l_2.
- Examples of n-point metric spaces that require distortion $\Omega(\log n)$ for embeddings into l_2 are so-called 'constant degree expander graphs' with the graph metric.
- Bourgain's thm. (below) shows that $\Omega(\log n)$ distortion requirement is the worst case.

Thm.: [Fréchet's embedding thm.]
Any countable metric space (X,d) admits an isometric embedding (i.e., one with distortion 1), into l_∞.

proof: For each $x \in X$ we define a coordinate $f_x : X \to \mathbb{R}$ via $f_x(u) := d(x,u)$ and $f : X \to l_\infty^{|X|}$ as $f(u) := (f_{x_1}(u), f_{x_2}(u), \dots)$. Then

$$\|f(u)-f(v)\|_\infty = \sup_{x\in X} |f_x(u) - f_x(v)| = \sup_{x \in X} |d(x,u) - d(x,v)| \begin{cases} \leq d(u,v) & \text{by triangle ineq.} \\ \geq d(u,v) & \text{by setting } x = u. \end{cases}$$ $\square$

Thm.: [Bourgain '85]

Let (X,d) be a metric space with $|X| =: n \in \mathbb{N}$. There exists a D-embedding into ℓ_p^k with $D = O(\log n)$ and $k = O((\log n)^2)$ for any $p \geq 1$.

proof (sketch): We will first consider the case $p=1$ and later show that this implies all other cases.

We will use a double index (i,j) for the coordinates where $i = 1, \dots, \log n$ and $j = 1, \dots, c_1 \log n$ with c_1 a suitable constant. For each (i,j) pick uniformly a random set $A_{ij} \subseteq X$ of size $|A_{ij}| = \frac{n}{2^i}$.

We define coordinates $f_{ij}(u) := \dfrac{d(u, A_{ij})}{k}$ as the minimal distance to the set A_{ij} rescaled by the number $k = c_1 \log^2 n$ of coordinates.

Then
$$\| f(u) - f(v) \|_1 = \frac{1}{k} \sum_{i,j} | d(u, A_{ij}) - d(v, A_{ij}) |$$
$$\overset{\Delta\text{ ineq.}}{\leq} \frac{1}{k} \sum_{i,j} d(u,v) = d(u,v).$$

So it remains to derive a lower bound.

Define $r_i := \min \{ r \in \mathbb{R} \mid |B_r(u)| = 2^i \wedge |B_r(v)| = 2^i \}$ for $i = 0, \dots, m$ where m is s.t. $\frac{1}{2} d(u,v) > r_m \geq c_2 d(u,v)$ for some constant $c_2 \in (0, \frac{1}{2})$.

Consider a given r_i with $|B_{r_i}(u)| = 2^i \wedge |B_{r_i}(v)| \geq 2^i$. For any j one can lower bound

$P[\, A_{ij} \cap B_{r_i}(u) = \emptyset \wedge A_{ij} \cap B_{r_{i-1}}(v) \neq \emptyset \,]$ by a non-zero constant.

In this case

$$f_{ij}(u) \geq \frac{r_i}{k} \quad \wedge \quad f_{ij}(v) \leq \frac{r_{i-1}}{k}$$ so that for fixed i

$$| f_{ij}(u) - f_{ij}(v) | \geq \frac{1}{k} (r_i - r_{i-1})$$ holds for a constant fraction of the $c_1 \log n$ j's. Consequently,

$$\| f(u) - f(v) \|_1 \geq \sum_i \frac{\Omega(\log n)}{k} (r_i - r_{i-1}) \underset{\substack{\uparrow \\ \text{telescopic sum with } r_0 = 0,\\ r_m \geq c_2 d(u,v)}}{=} \frac{d(u,v)}{O(\log n)}.$$

This shows the claim for $p=1$.

t.b.c. ...

For other values of p note that:

$$\| f(u) - f(v) \|_p = \left(\sum_{i,j} | f_{ij}(u) - f_{ij}(v) |^p \right)^{1/p}$$

$$\leq \left(\sum_{i,j} \frac{d(u,v)^p}{k^p} \right)^{1/p} = k^{-\frac{1}{q}} d(u,v) \quad \text{and}$$

$$\| f(u) - f(v) \|_p \, k^{1/q} \underset{\substack{\uparrow \\ \text{Hölder's ineq.}}}{\geq} \sum_{i,j} | f_{ij}(u) - f_{ij}(v) | \cdot 1 = \| f(u) - f(v) \|_1$$

$$\geq \frac{d(u,v)}{O(\log n)} .$$

$\square$

remarks:

- Suppose (Y, d) is a metric space with subset $X \subseteq Y$. Then the above embedding is such that $\| f(x) - f(y) \|_1 \leq d(x,y)$ for all $x, y \in Y$. Only the lower bound exploits that $x, y \in X$. Moreover, both only require a semi-metric space (where $d(x,y) = 0$ need not imply $x = y$).
- Note that the chosen embedding is independent of p.
- For $p = 2$ we can compose with the JL embedding and obtain $O(\log n)$ distortion into ℓ_2^k with $k = O(\log n)$.
- [Abraham, Bartal, Neiman 2006-2011] have shown that one can embed into any ℓ_p^k with $k = O(\log n)$ and distortion $O(\log n)$.
  They also prove that 'average distortion' of $O(1)$ can always be achieved in these cases.
- [Matousek '96] showed that one can embed into ℓ_p with distortion $O\left(\frac{\log n}{p}\right)$.
- One of the algorithmic applications of Bourgain's embedding thm. is an approximation algorithm for the 'sparsest cut problem' for graphs.

VII. Dvoretzky's theorem

Grothendieck asked whether every normed space $(\mathbb{R}^n, \|\cdot\|)$ has an almost Euclidean subspace of dim. k where $k \to \infty$ as $n \to \infty$. Dvoretzky ['61] answered this in the affirmative. Milman extended the result in the 70's.

Def.: (ε-net) Let (X,d) be a metric space and $\varepsilon > 0$. A subset $N \subseteq X$ is called an "ε-net" for X iff $\forall x \in X \; \exists y \in N : d(x,y) \leq \varepsilon$.

remarks:
- That is, N is an ε-net for X iff X can be covered by ε-balls around N.
- X is compact iff $|N| < \infty$.

Lemma: For $\varepsilon \in (0,1]$ there exists an ε-net N for S^{n-1} with cardinality $|N| \leq \left(1 + \frac{2}{\varepsilon}\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n = e^{n \log(3/\varepsilon)}$.

proof: Let N be a maximal ε-separated subset of S^{n-1}, i.e., $\forall x,y \in N: d(x,y) \geq \varepsilon$ and there is no $z \in S^{n-1}/N$ that we could add to N without violating this relation. Then N must be an ε-net (since otherwise there would be such a z) and the balls with centers in N and radius $\frac{\varepsilon}{2}$ are disjoint. Furthermore, all these balls are contained in $(1+\frac{\varepsilon}{2}) B_{l_2^n}$. Hence, $|N| \cdot \mathrm{vol}\left(\frac{\varepsilon}{2} B_{l_2^n}\right) \leq \mathrm{vol}\left(\left(1+\frac{\varepsilon}{2}\right) B_{l_2^n}\right)$ s.t.

$$|N| \left(\frac{\varepsilon}{2}\right)^n \leq \left(1+\frac{\varepsilon}{2}\right)^n \quad \text{and thus}$$
$$|N| \leq \left(1+\frac{2}{\varepsilon}\right)^n .$$

□

remarks:
- Note that the same argument (with the same bound) holds for an ε-net of $B_{l_2^n}$.
- One can derive a lower bound on $|N|$ that is exponential in n as well.

The following Lemma estimates how well one can approximate operator norms and condition numbers on an ε-net:

Lemma: Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be finite dimensional normed spaces, $T \in \mathcal{B}(X,Y)$ and $N \subseteq \partial B_X$ be an ε-net for $\partial B_X := \{x \in X \mid \|x\|_X = 1\}$.
If $\forall x \in N: \ a \leq \|Tx\|_Y \leq b$, then $\forall x \in X$:

$$\left(a - \frac{\varepsilon b}{1-\varepsilon}\right)\|x\|_X \leq \|Tx\|_Y \leq \frac{b}{1-\varepsilon}\|x\|_X .$$

proof: Since $\forall x \in \partial B_X \ \exists y \in N: \|x-y\| \leq \varepsilon$ we have
$\|Tx\| \leq \|Ty\| + \|T(x-y)\| \leq b + \|T\|\varepsilon$ so that $\|T\| \leq \frac{b}{1-\varepsilon}$ proving the right inequality.
For the left one, note that $\forall x \in \partial B_X$:
$\|Tx\| \geq \|Ty\| - \|T(x-y)\| \geq a - \|T\|\varepsilon \geq a - \frac{\varepsilon b}{1-\varepsilon}$. $\square$

Thm.: (Dvoretzky criterion)

Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space with $L := \sup_{x \in S^{n-1}} \|x\|$ and m a median of the map $S^{n-1} \ni x \mapsto \|x\|$ w.r.t. the spherical measure. For any $\varepsilon \in (0,1)$ a Haar-random subspace V of dim. k satisfies:

$$\forall x \in V: \ \frac{1-3\varepsilon}{1-\varepsilon} m \|x\|_2 \leq \|x\| \leq \frac{1+\varepsilon}{1-\varepsilon} m \|x\|_2$$

with non-zero probability if $k < \frac{\varepsilon^2}{\log(3/\varepsilon)} \frac{n m^2}{2L^2}$.

<u>proof:</u> Let $V_0 \subseteq \mathbb{R}^n$ be a subspace of dimension k (to be chosen later) and $\mathcal{N} \subseteq V_0 \cap S^{n-1} \cong S^{k-1}$ be an ε-net with $|\mathcal{N}| \leq e^{k \log(3/\varepsilon)}$.

Denote the Haar-measure on $O(n)$ by μ and fix $x_0 \in S^{n-1} \cap V_0$. Then

$$\mu\left\{ U \in O(n) \,\middle|\, \big| \|Ux_0\| - m \big| > \varepsilon m \right\} = \sigma\left\{ x \in S^{n-1} \,\middle|\, \big| \|x\| - m \big| > \varepsilon m \right\}$$

$$\leq e^{-n\varepsilon^2 m^2/(2L^2)}$$

| Lévy's Lemma with optimal constants

by Lévy's Lemma since $x \mapsto \|x\|$ is L-Lipschitz.

Now vary x_0 over the ε-net. Then $(1-\varepsilon)m \leq \|Ux_0\| \leq (1+\varepsilon)m$ holds for all $x_0 \in \mathcal{N}$ with probability at least

$$1 - |\mathcal{N}|\, e^{-n\varepsilon^2 m^2/(2L^2)} \geq 1 - \exp\left[k \log\left(\frac{3}{\varepsilon}\right) - \frac{n\varepsilon^2 m^2}{2L^2} \right],$$

which is larger than zero if

$$k < \frac{\varepsilon^2}{\log(3/\varepsilon)} \frac{n m^2}{2L^2}.$$

In this case there is a $U \in O(n)$ s.t. with $V := UV_0$ we have $\big| \|x\| - m \big| \leq \varepsilon m$ holds for all $x \in U\mathcal{N} \subseteq V \cap S^{n-1}$.

Applying the approximation Lemma to T being the identity operator $T: (V, \|\cdot\|_2) \to (V, \|\cdot\|)$ we obtain:

$$\forall x \in V: \qquad \frac{1-3\varepsilon}{1-\varepsilon} m \|x\|_2 \leq \|x\| \leq \frac{1+\varepsilon}{1-\varepsilon} m \|x\|_2 .$$

$\square$

Cor.: Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space and $\varepsilon \in (0,1)$. Then there exists a k-dim. subspace $V \subseteq X$ s.t. $d(V, l_2^k) \leq \frac{1+\varepsilon}{1-3\varepsilon}$ if $k < \frac{\varepsilon^2}{\log(3/\varepsilon)} \frac{n}{2} d(X, l_2^n)^{-2}$.

proof: Suppose $d = d(X, l_2^n)$, i.e. $\exists T \in \mathcal{B}(\mathbb{R}^n)$, $a, b \in \mathbb{R}_{>0}$ s.t. $\forall x \in X$:
$a\|x\|_2 \leq \|Tx\| \leq b\|x\|_2$ and $\frac{b}{a} = d$.
We apply the Dvoretzky criterion to the norm $x \mapsto \|Tx\|$. Then $b = L$ is the corresponding Lipschitz constant and since $\forall x \in S^{n-1}$: $\|Tx\| \geq a$, we have $m \geq a$. Consequently, $\frac{m}{L} \geq \frac{a}{b} = \frac{1}{d}$ and the Banach-Mazur distance between l_2^k and the subspace V is upper bounded by $\left(\frac{1+\varepsilon}{1-\varepsilon} m\right) / \left(\frac{1-3\varepsilon}{1-\varepsilon} m\right) = \frac{1+\varepsilon}{1-3\varepsilon}$. □

remark:

- This shows that there are almost Euclidean subspaces of dim. $\Omega\left(\frac{n}{d(X, l_2^n)^2}\right)$, which is a non-trivial statement unless $d(X, l_2^n)$ has the worst-case scaling, namely $\sim \sqrt{n}$ (like l_∞^n, see John's thm. on p. 11).
- If $X = l_p^n$ with $p \in (2, \infty)$ this means that almost Euclidean subspaces of dimension $\Omega(n^{2/p})$ exist. If $p \in [1,2)$ this holds even with dimension $\Omega(n)$. ($\rightarrow$ Exercise)
- In order to get a more general bound we have to lower bound $\frac{m}{L}$ in the Dvoretzky criterion.
- Geometrically speaking, the above proof guarantees sections of B_X that are almost ellipsoids. The following implies that even spherical ellipsoids can always be obtained.

Lemma: Let $k \in \mathbb{N}$ and $A \in \mathbb{R}^{(2k-1)\times(2k-1)}$ be a symmetric matrix. Then there exists $\lambda_0 \in \mathbb{R}$ and an orthogonal projection P with $\mathrm{tr}[P] = k$ s.t. $PAP = \lambda_0 P$.

<u>proof:</u> Let $\lambda_{1-k} \leq \dots \leq \lambda_0 \leq \dots \leq \lambda_{k-1}$ be the eigenvalues of A and $\{x_i\}_{i=1-k}^{k-1}$ the corresponding orthonormal eigenvectors. Set $y_0 := x_0$ and for $i = 1, \dots, k-1$: $y_i := x_{-i} \cos\varphi_i + x_i \sin\varphi_i$. Then $\{y_i\}_{i=0}^{k-1}$ will be orthonormal and $\langle y_i, A y_j \rangle = \delta_{ij} \left(\lambda_{-i} \cos^2\varphi_i + \lambda_i (1-\cos^2\varphi_i) \right)$. Hence, we can always choose the φ_i's s.t. $\langle y_i, A y_j \rangle = \delta_{ij} \lambda_0$. Then $P = \sum_{i=0}^{k-1} y_i y_i^*$ is the sought projector. □

<u>Cor.:</u> <u>(spherical sections of ellipsoids)</u> If $\mathcal{E} \in \mathbb{R}^n$ is an ellipsoid, then there is a subspace V of dim. $k \geq \frac{n}{2}$ s.t. for some $\lambda_0 > 0$:
$\lambda_0 \mathcal{E} \cap V = B_{l_2^k}$.

<u>proof:</u> W.l.o.g. assume that $n = 2k-1$ and $\mathcal{E} = T B_{l_2^n} = \{ x \in \mathbb{R}^n \mid \langle x, \underbrace{(T^{-1})^*(T^{-1})}_{=:A} x \rangle \leq 1 \}$.
Apply the Lemma and set $V = P\mathbb{R}^n$. Then $\mathcal{E} \cap V = \{ x \in \mathbb{R}^n \mid x = Px \wedge \langle x, Ax \rangle \leq 1 \} = \lambda_0^{-1} B_{l_2^k}$. Moreover, $\lambda_0 > 0$ since $A > 0$. □

<u>Lemma:</u> <u>(Dvoretzky-Rogers Lemma ['50])</u>
Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space with unit ball B_X s.t. the max. volume ellipsoid in B_X is $B_{l_2^n} \subseteq B_X$. Then there is an l_2^n-orthonormal basis $\{x_i\}_{i=1}^n$ s.t. $\|x_i\| \geq 1 - \frac{i}{n}$ $\forall i$.
(In particular, $\|x_l\| \geq \frac{1}{2} \|x_l\|_2$ $\forall l = 1, \dots, \lfloor \frac{n}{2} \rfloor$)

<u>proof:</u> Let $P_k \in \mathcal{B}(l_2^n)$ be a k-dim. orth projector, $T: l_2^n \to X$ the identity map and $\|\cdot\|$ on $\mathcal{B}(l_2^n, X)$ the operator norm. Then

$$1 - \frac{k}{n} = \frac{1}{n} \operatorname{tr}[\mathbb{1} - P_k] = \frac{1}{n} \operatorname{tr}[(T^{-1}T)(\mathbb{1} - P_k)]$$

$$\leq \frac{1}{n} \|T^{-1}\|^* \|T(\mathbb{1} - P_k)\| \quad \text{since} \quad \|T^{-1}\|^* := \sup_{\|S\|=1} |\operatorname{tr}[T^{-1}S]|$$

$$= \|T(\mathbb{1} - P_k)\| \quad \text{by Lewis' thm. (see p. 7)}$$

Consequently, there is an x s.t. $\|(\mathbb{1} - P_k)x\| \geq (1 - \frac{k}{n}) \|x\|_2 \geq (1 - \frac{k}{n}) \|(\mathbb{1} - P_k)x\|_2$

and with $x_k := (\mathbb{1} - P_k)x / \|(\mathbb{1} - P_k)x\|_2$ we get $\|x_k\| \geq (1 - \frac{k}{n})$.
Beginning with any one-dim. orth. proj. P_1 and setting $P_{k+1} := P_k + x_k x_k^*$ we obtain the sought ONB. □

Proposition: Let $n > 7$ and $X = (\mathbb{R}^n, \|\cdot\|)$ a normed space s.t. $B_{L_2^n}$ is the maximum volume ellipsoid inside B_X. Then:

$$\int_{S^{n-1}} \|x\| \, d\sigma(x) \geq \frac{1}{6}\sqrt{\frac{\log n}{n}} .$$

proof: Let $\{x_i\}_{i=1}^n$ be the ONB from the Dvoretzky-Rogers Lemma, i.e. $\|x_i\| \geq \frac{1}{2} \ \forall i \leq \frac{n}{2}$

$$\int_{S^{n-1}} \|x\| \, d\sigma(x) = \int_{S^{n-1}} \Big\| \sum_{i=1}^n a_i x_i \Big\| \, d\sigma(a)$$

invariance of σ under $O(n)$

$$= \frac{1}{2}\int_{S^{n-1}} \left(\Big\| \sum_{i=1}^{n-1} a_i x_i + a_n x_n \Big\| + \Big\| \sum_{i=1}^{n-1} a_i x_i - a_n x_n \Big\| \right) d\sigma(a)$$

$\|x\| \leq \frac{1}{2}(\|x+y\| + \|x-y\|)$

$$\geq \int_{S^{n-1}} \max\left\{ \Big\| \sum_{i=1}^{n-1} a_i x_i \Big\|, \|a_n x_n\| \right\} d\sigma(a)$$

repeated application of previous steps

$$\geq \dots \geq \int_{S^{n-1}} \max\{\|a_i x_i\|\}_{i=1}^n \, d\sigma(a)$$

Dvoretzky-Rogers

$$\geq \frac{1}{2}\int_{S^{n-1}} \max\{|a_i|\}_{i \leq \frac{n}{2}} \, d\sigma(a)$$

Let γ_n be the standard Gaussian prob. measure in $\mathbb{R}^n$, i.e., $\gamma_n(A) = (2\pi)^{-n/2}\int_A e^{-\|x\|^2/2} dx$. If a random vector $g \in \mathbb{R}^n$ is distributed according to γ_n (i.e., $g_1, \dots, g_n$ are i.i.d. $N(0,1)$ random variables), then due to $O(n)$-invariance of γ and the uniqueness of the spherical measure (as Haar measure on S^{n-1}), $a := \frac{g}{\|g\|_2}$ is distributed according to σ. Hence,

$$\int_{S^{n-1}} \max\{|a_i|\}_{i=1}^{\lfloor \frac{n}{2} \rfloor} d\sigma(a) = \mathbb{E}\left(\frac{\max\{|g_i|\}_{i=1}^{\lfloor \frac{n}{2} \rfloor}}{\|g\|_2} \right) = \frac{\mathbb{E}\left(\max\{|g_i|\}_{i=1}^{\lfloor \frac{n}{2} \rfloor}\right)}{\mathbb{E}(\|g\|_2)} ,$$

where $\mathbb{E}$ is w.r.t. γ_n.

Since the random vector $\frac{g}{\|g\|_2} =: a$ and the random variable $\|g\|_2$ are independent, we have for every $f : \mathbb{R}^n \to \mathbb{R}$ that

$$\mathbb{E}(f(a) \cdot \|g\|_2) = \mathbb{E}(f(a)) \cdot \mathbb{E}(\|g\|_2)$$

By Jensen's ineq. we can bound $\mathbb{E}(\|g\|_2) \leq \left(\mathbb{E}\sum_{i=1}^n |g_i|^2\right)^{1/2} = \sqrt{n}$.

Bounding the enumerator requires more work:

$$\mathbb{E}\left(\max\{|g_i|\}_{i\leq\frac{n}{2}}\right) = \int_{\mathbb{R}^n} \max\{|g_i|\}_{i\leq\frac{n}{2}}\, d\gamma_n(g)$$

$$\geq t\ \gamma_n\left\{g\in\mathbb{R}^n \,\middle|\, \max\{|g_i|\}_{i\leq\frac{n}{2}} \geq t\right\}, \quad \text{later set } t=\sqrt{2\log n} > \sqrt{2} \ (\text{so } n>2)$$

$$= t\left[1-\left(\gamma_1\{g\in\mathbb{R} \mid |g|<t\}\right)^{\frac{n}{2}}\right]$$

$$= t\left[1-\left(1-\gamma_1\{g\in\mathbb{R} \mid |g|\geq t\}\right)^{\frac{n}{2}}\right]$$

$$= t\left[1-\left(1-\frac{2}{\sqrt{2\pi}}\int_t^\infty e^{-x^2/2}\,dx\right)^{\frac{n}{2}}\right]$$

$$\left\{\begin{aligned} \int_t^\infty e^{-\frac{x^2}{2}}dx &\geq \int_t^\infty (1-3x^{-4})\,e^{-\frac{x^2}{2}}dx \\ &= (t^{-1}-t^{-3})\,e^{-t^2/2} \\ &\geq \frac{1}{2t}\,e^{-t^2/2} \quad \text{if } t>\sqrt{2} \\ &\qquad (\text{i.e. } n>7) \end{aligned}\right.$$

$$\overset{t:=\sqrt{\log n}}{\geq} \sqrt{\log n}\left[1-\left(1-(2\pi n\log n)^{-\frac{1}{2}}\right)^{\frac{n}{2}}\right]$$

$$\geq \frac{1}{3}\sqrt{\log n} \qquad \text{for } n>7.$$

Putting things together we obtain

$$\int_{S^{n-1}} \|x\|\, d\sigma(x) \geq \frac{1}{2}\,\frac{\mathbb{E}\left(\max\{|g_i|\}_{i\leq\frac{n}{2}}\right)}{\mathbb{E}(\|g\|_2)} \geq \frac{1}{6}\sqrt{\frac{\log n}{n}}. \qquad \square$$

Replacing the median in the Dvoretzky criterion by the average and using the above estimate we finally obtain:

Thm.: (Dvoretzky's thm. - Milman's version)

$\forall \varepsilon>0\ \exists c>0$ s.t. for every normed space $X=(\mathbb{R}^n,\|\cdot\|)$ there is a subspace V of dimension $k\geq c\log n$ and $s>0$ s.t. $\forall x\in V$:

$$(1-\varepsilon)\|x\|_2 \leq s\|x\| \leq (1+\varepsilon)\|x\|_2.$$

remarks:

- We can achieve $c(\varepsilon)=\Omega(\varepsilon^2)$
- This implies that if X is an infinite-dim. real Banach space, then $\forall n\in\mathbb{N}\ \forall\varepsilon>0$ there is a subspace V s.t. $d(V,\ell_2^n)\leq 1+\varepsilon$.

- In the proof (p.32) we see that the cube (i.e. ℓ_∞^n) is the worst case. In fact, for ℓ_∞^n an almost Euclidean subspace of dim $\sim \log n$ is best possible. That is, the $\log n$-scaling in Dvoretzky's thm. is e.g. optimal.

Lemma: Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space and $X^* = (\mathbb{R}^n, \|\cdot\|_*)$ its dual. With $M := \int_{S^{n-1}} \|x\| \, d\sigma(x)$, $M^* := \int_{S^{n-1}} \|x\|_* \, d\sigma(x)$ it holds that $MM^* \geq 1$.

proof:

$$MM^* \underset{\text{Hölder}}{\geq} \left(\int_{S^{n-1}} \|x\|_*^{1/2} \|x\|^{1/2} \, d\sigma(x) \right)^2$$

$$\geq \left(\int_{S^{n-1}} |\langle x, x \rangle|^{1/2} \, d\sigma(x) \right)^2 = 1$$

$\|x\|_*$ is the operator norm of the functional $X \ni y \mapsto \langle x, y \rangle$. $\square$

Def.: (Dvoretzky dimension) Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space. We call the "<u>Dvoretzky dimension</u>" $k(X)$ the largest integer s.t. there is a subspace $V \subseteq X$ of dimension $k(X)$ that satisfies $d(X, \ell_2^{k(X)}) \leq 1+\varepsilon$.

Clearly, $k(X)$ depends on ε.

Prop.: $\forall \varepsilon > 0 \; \exists c > 0$, s.t. the Dvoretzky dimensions of a normed space $X = (\mathbb{R}^n, \|\cdot\|)$ and its dual X^* satisfy:

$$k(X)\, k(X^*) \geq \frac{c n^2}{d(X, \ell_2^n)^2} .$$

proof: Suppose $a\|x\|_2 \leq \|x\| \leq b\|x\|_2$ for some $a, b > 0$ s.t. $d(X, \ell_2^n) = \frac{b}{a}$.

Then $\frac{1}{b}\|x\|_2 \leq \|x\|_* \leq \frac{1}{a}\|x\|_2$ since $\|x\|_* := \sup\{ |\langle y, x \rangle| \mid \|y\| = 1 \}$

From Dvoretzky's thm. we know $k(X) \geq c(\varepsilon)\, n \left(\frac{M}{b}\right)^2$ and

$k(X^*) \geq c(\varepsilon)\, n \left(\frac{M^*}{1/a}\right)^2$.

Consequently, $k(X)\,k(X^*) \geq c(\varepsilon)^2 n^2 \left(\frac{MM^*}{b/a}\right)^2 \underset{\text{Lemma}}{\geq} \frac{c n^2}{d(X, \ell_2^n)^2}$. $\square$

Lemma: (Duality between projections and sections)
Let $K \subseteq \mathbb{R}^n$ be a convex body with $0 \in K$ and $S \subseteq \mathbb{R}^n$ a subspace with corresponding orth. proj. P_S. Then $S \cap (K \cap S)^\circ = P_S(K^\circ)$.

proof: Consider

$$S \cap (P_S(K^\circ))^\circ = \{ x \in S \mid \forall y \in P_S(K^\circ) : \underbrace{\langle x, y \rangle}_{= \langle P_S x, P_S y \rangle = \langle x, y \rangle} \leq 1 \}$$

$$= \{ x \in S \mid \forall y \in K^\circ : \langle x, y \rangle \leq 1 \}$$

$$= \{ x \in \mathbb{R}^n \mid \forall y \in K^\circ : \langle x, y \rangle \leq 1 \} \cap S \underset{\text{Bipolar thm.}}{=} K \cap S$$

Applying the bipolar thm. again, but in S rather than in $\mathbb{R}^n$, we obtain
$(K \cap S)^\circ \cap S = P_S(K^\circ)$. $\square$

Cor.: (Dvoretzky with projections)
$\forall \varepsilon > 0$ there is a $c > 0$ s.t. for any symmetric convex body $K \subseteq \mathbb{R}^n$ there is an orthogonal projections P onto a space of dim. $d \geq c \log n$ s.t.
$d(P(K), B_{\ell_2^d}) \leq 1 + \varepsilon$.

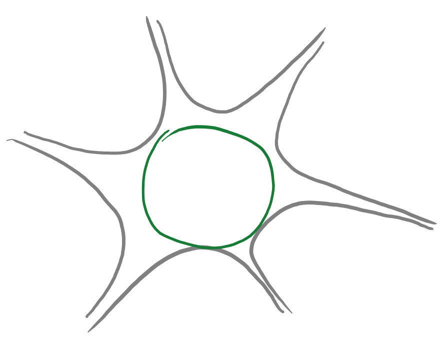
proof: By Dvoretzky's thm. there is a subspace S with $\dim(S) =: d \geq c \log n$ s.t.

$$1 + \varepsilon \geq d(K^\circ \cap S, \ell_2^d) = d(K^\circ \cap S, B_{\ell_2^d}) \underset{\text{take the polar in } S}{=} d((K^\circ \cap S)^\circ \cap S, B_{\ell_2^d}^\circ) =$$

$$= d(P_S(K), B_{\ell_2^d}).$$

$\square$

Recall Milman's "hyperbolic drawing" of high-dimensional convex sets and the statement that intersecting or averaging randomly rotated versions of such a set becomes almost spherical. This is the "global version" of Dvoretzky's thm.:



Thm.: (Dvoretzky - "global version")

$\forall \eta > 0$ there is a $c > 0$ s.t. for any normed space $X = (\mathbb{R}^n, \|\cdot\|)$ there are orthogonal transformations $U_1, \dots, U_t$ with $t \le c\left(\frac{L}{M}\right)^2$, where $M := \int_{S^{n-1}} \|x\| \, d\sigma(x)$, $L := \sup_{x \in S^{n-1}} \|x\|$ s.t.

$$\forall z \in S^{n-1}: \quad (1-\eta) M \|z\|_2 \le \frac{1}{t} \sum_{i=1}^{t} \|U_i z\| \le (1+\eta) M \|z\|_2 .$$

proof (sketch): For $x = (x_1, \dots, x_t) \in (S^{n-1})^t$ define $\nu(x) := \frac{1}{t}\sum_{i=1}^{t} \|x_i\|$. Then $\forall x, y \in (S^{n-1})^t$:

$$|\nu(x) - \nu(y)| \le \frac{1}{t} \sum_{i=1}^{t} \|x_i - y_i\|$$
$$\le \frac{1}{\sqrt{t}} \left(\sum_{i=1}^{t} \|x_i - y_i\|^2\right)^{1/2} \qquad \text{since } \|z\|_1 \le \sqrt{t}\,\|z\|_2 \text{ for } z \in \mathbb{R}^t$$
$$\le \underbrace{\frac{L}{\sqrt{t}}}_{=: L_\nu} \|x - y\|_2 \qquad \text{since } \|z\| \le L \|z\|_2 \ \forall z \in \mathbb{R}^n$$

From a variant of Levy's Lemma for products of spheres one obtains with $m := \mathrm{med}(\nu)$ w.r.t. the product spherical measure $\sigma \times \dots \times \sigma$:

$$P\left[\, |\nu(x) - m| > \delta m \,\right] \le \exp\left[-c\left(\frac{\delta m}{L_\nu}\right)^2 n\right]$$

for some constant $c > 0$ and any $\delta \in (0,1)$. Equivalently, for a fixed $z \in S^{n-1}$:

$$(1-\delta) m \le \frac{1}{t}\sum_{i=1}^{t} \|U_i z\| \le (1+\delta) m \quad (*)$$

only fails for $U \in O(n)^t$ in a subset of measure smaller than $\exp\left[-c\left(\frac{\delta m}{L_\nu}\right)^2 n\right]$. If N is an ε-net for S^{n-1} with $|N| \le e^{n \log\left(\frac{3}{\varepsilon}\right)}$, then $(*)$ fails for at least one $z \in N$ with prob. smaller than

$$N \cdot \exp[\dots] = \exp\left[n\left(\log\left(\tfrac{3}{\varepsilon}\right) - c\left(\tfrac{\delta m}{L}\right)^2 t\right)\right],$$

which is in turn smaller than 1 if $t \ge \frac{1}{c}\log\left(\frac{3}{\varepsilon}\right)\left(\frac{L}{\delta m}\right)^2$.

Choosing ε sufficiently small we obtain $(*)$ for all $z \in S^{n-1}$ (at the cost of slightly increasing δ) so that with $\left|m - \int_{S^{n-1}} \|x\| \, d\sigma(x)\right| = O\left(\frac{1}{\sqrt{n}}\right)$ the result follows. □

remark: From our previous considerations we know that

$$\left(\frac{L}{M}\right) \le d(X, \ell_2^n) \quad \text{and} \quad \left(\frac{L}{M}\right)^2 \le c\,\frac{n}{\log n} \quad \text{in general.}$$