Modern Approximation Theory

. . . draft script. . .

Contents

1 Introduction

1.1 Banach spaces

Definition 1. Let X be a vector space over $\mathbb{R}$ (or $\mathbb{C}$) equipped with a *norm* $||\cdot|| : X \to [0,\infty)$. It means, that

- There are operations $+ : X \times X \to X$ and $\cdot : \mathbb{R} \times X \to X$ with the usual properties.
- The function $||\cdot||$ satisfies
  (1) $||x|| = 0$ if, and only if,[1] $x = 0$,
  (2) $||\alpha x|| = |\alpha| \cdot ||x||$ for all $\alpha \in \mathbb{R}$ (or all $\alpha \in \mathbb{C}$) and all $x \in X$,
  (3) $||x + y|| \leq ||x|| + ||y||$ for all $x, y \in X$.

If X is complete[2], it is called a *Banach space*.

Remark 1. We shall write $||x||_X$ or $||x|X||$ to emphasise the space X in the notation.

Remark 2. We recall the most classical Banach spaces

- $\mathbb{R}^n, \mathbb{C}^n, \ell_p^n(\mathbb{R}), \ell_p^n(\mathbb{C})$ with $1 \leq p \leq \infty$,
- ℓ_p, c_0, c_{00} with $1 \leq p \leq \infty$,
- $L_p([0,1]), L_p(\Omega), C(\Omega)$ with $1 \leq p \leq \infty$.

Remark 3. The most important concepts of functional analysis are

- completeness,
- duality,
- convexity,
- compactness.

Completeness shall be used only very rarely in this text.

1.2 Duality and convexity

Definition 2. Let X, Y be two Banach spaces.

- $f : X \to Y$ is *linear*, iff
$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$
  for all $\alpha, \beta \in \mathbb{R}$ (or $\mathbb{C}$) and all $x, y \in X$.
- If $f : X \to Y$ is linear, it is *continuous*, iff there exists a real number $M > 0$, such that
$$||f(x)||_Y \leq M \, ||x||_X$$
  for all $x \in X$.
- If $Y = \mathbb{R}$ (or $Y = \mathbb{C}$) and $f : X \to Y$ is linear and continuous, then it is called *functional*.

[1] This shall be sometimes abbreviated as "iff" for short
[2] i.e. every Cauchy sequence is convergent

- $\mathcal{L}(X,Y) = \{f : X \to Y, f \text{ is linear and continuous}\}$.
- $X' = \mathcal{L}(X,\mathbb{R})$ or $X' = \mathcal{L}(X,\mathbb{C})$, respectively, is the *dual space* of X.

Remark 4.

- $\mathcal{L}(X,Y)$ is complete with respect to the *operator norm*

$$||f|\mathcal{L}(X,Y)|| = \sup_{x\neq 0} \frac{||f(x)||_Y}{||x||_X} = \sup_{x\in X:||x||=1} ||f(x)||_Y.$$

Especially,

$$||f|X'|| = \sup_{x\neq 0} \frac{|f(x)|}{||x||_X} = \sup_{x\in X:||x||=1} |f(x)|.$$

- The supremum above may not be attained, cf. Exercise 2.
- It is usual to identify the dual space with some other well known space. For example, $(\ell_1)' = \ell_\infty$, cf. Exercise 3. In this sense, we have

$$(\ell_p)' = \ell_{p'}, \quad \text{where} \quad \frac{1}{p} + \frac{1}{p'} = 1, \quad 1 \le p < \infty.$$

Let us recall, that $(\ell_\infty^m)' = \ell_1^m$, but $(\ell_\infty)' \neq \ell_1$, cf. Exercise 8.

- The operators may be composed in the usual way, i.e.

$$S \in \mathcal{L}(X,Y), \quad T \in \mathcal{L}(Y,Z) \quad \Longrightarrow T\circ S \in \mathcal{L}(X,Z)$$

and

$$||T\circ S|\mathcal{L}(X,Z)|| \le ||T|\mathcal{L}(Y,Z)|| \cdot ||S|\mathcal{L}(X,Y)||.$$

Definition 3. Let X be a vector space.

a) A set $M \subset X$ is called *convex*, iff

$$\forall x,y \in M \quad \forall \lambda : 0 \le \lambda \le 1 \quad \lambda x + (1-\lambda)y \in M.$$

b) Let $M \subset X$ be convex and let $f : M \to \mathbb{R}$ be an arbitrary function. Then f is *convex*, iff

$$\forall x,y \in M \quad \forall \lambda : 0 \le \lambda \le 1 \quad f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y).$$

c) Let $M \subset X$. Then we define the *convex hull* of M by

$$\text{conv } M = \bigcap_{\substack{K\supset M \\ K \text{ convex}}} K.$$

Remark 5.

- Convexity is defined only algebraically. No topology or norm is assumed on X and no continuity of f is necessary.
- Let X be a Banach space and let

$$B_X = \{x \in X : ||x||_X \le 1\}$$

be its closed *unit ball*. Then B_X is convex.

Definition 4. a) Let X be a vector space and let $||\cdot|| : X \to [0,\infty)$ with
(1) $||x|| = 0$ if, and only if, $x = 0$,
(2) $||\alpha x|| = |\alpha| \cdot ||x||$ for all $\alpha \in \mathbb{R}$ (or all $\alpha \in \mathbb{C}$) and all $x \in X$,
(3) $||x + y|| \le C(||x|| + ||y||)$ for some $C \ge 1$ and all $x, y \in X$.

Then $||\cdot||$ is called a *quasi-norm* and X is a *quasi-normed space*. If X is even complete in this quasi-norm, it is called also a *quasi-Banach space*.

b) Let X be a vector space and let $||\cdot|| : X \to [0,\infty)$ with
(1) $||x|| = 0$ if, and only if, $x = 0$,
(2) $||\alpha x|| = |\alpha| \cdot ||x||$ for all $\alpha \in \mathbb{R}$ (or all $\alpha \in \mathbb{C}$) and all $x \in X$,
(3) $||x + y||^p \le ||x||^p + ||y||^p$ for some $0 < p \le 1$ and all $x, y \in X$.

Then $||\cdot||$ is called a *p-norm* and X is a *p-normed space*. If X is even complete in this p-norm, it is called also a *p-Banach space*.

Example 1. Let us consider the space $L_p([0,1])$ with $0 < p < 1$. This is a space of (equivalence classes of) measurable functions on $[0,1]$, equipped with a mapping

$$||f|L_p([0,1])|| = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}.$$

- Then it holds

$$\begin{aligned} ||f+g|L_p([0,1])||^p &= \int_0^1 |f(x)+g(x)|^p dx \le \int_0^1 |f(x)|^p + |g(x)|^p dx \\ &= ||f|L_p([0,1])||^p + ||g|L_p([0,1])||^p. \end{aligned}$$

  Here, we used the first inequality from Exercise 5a) with $a = |f(x)|$ and $b = |g(x)|$. Hence, $||\cdot||$ is a p-norm.

- On the other hand, we get

$$\begin{aligned} ||f+g|L_p([0,1])|| &= \left(\int_0^1 |f(x)+g(x)|^p dx\right)^{1/p} \le \left(\int_0^1 |f(x)|^p + |g(x)|^p dx\right)^{1/p} \\ &= \left(\int_0^1 |f(x)|^p dx + \int_0^1 |g(x)|^p dx\right)^{1/p} = \big(||f|L_p([0,1])||^p + ||g|L_p([0,1])||^p\big)^{1/p} \\ &\le 2^{1/p-1}\big(||f|L_p([0,1])|| + ||g|L_p([0,1])||\big). \end{aligned}$$

  Here, we have used the second inequality from Exercise 5b) with $a = ||f|L_p([0,1])||$ and $b = ||g|L_p([0,1])||$. Hence $||\cdot||$ is also a quasi-norm.

Next theorem shows, that this behaviour is no coincidence.

Theorem 5. *a) Let X be a p-Banach space. Then X is also a quasi-Banach space with $C = 2^{1/p-1}$.*

b) Let X be a quasi-Banach space with the quasi-norm $||\cdot||$. Then there is a $0 < p \le 1$ and a p-norm $||\cdot||_p$, which is equivalent [3] to $||\cdot||$.

[3] This means, that there are constants c_1 and c_2 such that $c_1||x|| \le ||x||_p \le c_2||x||$ for all $x \in X$. The topologies induced by equivalent quasi-norms are identical.

Proof. The proof of a) copies the proof given in Example 1.

$$||f+g|| \le \big(||f||^p + ||g||^p\big)^{1/p} \le [2^{1-p}(||f||+||g||)^p]^{1/p} = 2^{1/p-1}\big(||f||+||g||\big).$$

The proof of b) is the essential part.

Let C denote the constant from the triangle inequality for $||\cdot||$. We put $C_0 = 2C \ge 2$.

We define p through $C_0^p = 2$ (i.e. $0 < p \le 1$) and put

$$||f||_p = \inf_{f=f_1+\cdots+f_m} \big(||f_1||^p + \cdots + ||f_m||^p\big)^{1/p}.$$

The infimum runs over all decompositions $f = f_1 + \cdots + f_m$.

It remains to prove, that $||\cdot||_p$ has all the properties of a p-norm. This is unfortunately a bit technical. We start with the following observation.

Step 1. We prove through induction, that

$$||f_1 + \cdots + f_m|| \le \max_{1\le j\le m} (C_0^j ||f_j||). \tag{1.1}$$

- The case $m = 2$ is trivial.

$$||f_1+f_2|| \le C(||f_1||+||f_2||) \le \max(C_0||f_1||, C_0||f_2||) \le \max(C_0||f_1||, C_0^2||f_2||).$$

- The step $m-1 \to m$ follows.

$$\begin{aligned} ||f_1+\ldots+f_m|| &\le C(||f_1|| + ||f_2+\cdots+f_m||) \le \max(C_0||f_1||, C_0||f_2+\cdots+f_m||) \\ &\le \max(C_0||f_1||, C_0\max(C_0||f_2||, C_0^2||f_3||, \ldots, C_0^{m-1}||f_m||)) \\ &= \max(C_0||f_1||, C_0^2||f_2||, C_0^3||f_3||, \ldots, C_0^m||f_m||)). \end{aligned}$$

Step 2. We show, that $||\cdot||_p^p$ is subadditiv, i.e.

$$||f+g||_p^p \le ||f||_p^p + ||g||_p^p.$$

Let $f = f_1 + \cdots + f_m$ and $g = g_1 + \cdots + g_n$, such that

$$||f||_p \ge (1-\varepsilon)(||f_1||^p + \cdots + ||f_m||^p)^{1/p} \quad \text{and} \quad ||g||_p \ge (1-\varepsilon)(||g_1||^p + \cdots + ||g_n||^p)^{1/p},$$

where $\varepsilon > 0$ is arbitrary. Then $f+g = f_1+\cdots+f_m+g_1+\cdots+g_n$ and

$$||f+g||_p \le (||f_1||^p + \cdots + ||f_m||^p + ||g_1||^p + \cdots + ||g_n||^p)^{1/p} \le \left[\left(\frac{||f||_p}{1-\varepsilon}\right)^p + \left(\frac{||g||_p}{1-\varepsilon}\right)^p\right]^{1/p}.$$

Hence

$$||f+g||_p^p \le \frac{1}{(1-\varepsilon)^p}\big(||f||_p^p + ||g||_p^p\big)$$

and we let $\varepsilon \to 0$.

It follows trivially, that $||\alpha f||_p = |\alpha| \cdot ||f||_p$ and $||f||_p \le ||f||$. We show, that there is a constant $A \ge 1$, such that

$$||f|| \le A||f||_p, \quad f \in X. \tag{1.2}$$

This inequality proves, that $||\cdot||$ and $||\cdot||_p$ are equivalent and gives as a byproduct, that $||f||_p = 0$ if, and only if, $f = 0$.

Step 3. Proof of (1.2).

We show by induction, that

$$||f_1 + \cdots + f_m|| \le C_0(N(f_1)^p + \cdots + N(f_m)^p)^{1/p}, \tag{1.3}$$

where

$$N(f) = \begin{cases} 0, & \text{if} \quad f = 0, \\ C_0^k, & \text{if} \quad C_0^{k-1} < ||f|| \le C_0^k, \quad k \in \mathbb{Z}. \end{cases}$$

From (1.3) follows (1.2) immediately. If $f = f_1 + \cdots + f_m$, then

$$\begin{aligned} ||f|| = ||f_1 + \cdots + f_m|| &\le C_0(N(f_1)^p + \cdots + N(f_m)^p)^{1/p} \\ &\le C_0((C_0||f_1||)^p + \cdots + (C_0||f_m||)^p)^{1/p} = C_0^2(||f_1||^p + \cdots + ||f_m||^p)^{1/p} \end{aligned}$$

and we may consider the infimum over $f = f_1 + \cdots + f_m$.

- The proof of (1.3) for $m = 1$ is simple:

$$||f_1|| \le N(f_1) \le C_0(N(f_1)^p)^{1/p}.$$

- The proof of $m \to m+1$ in (1.3).

  We assume, that $||f_1|| \ge \cdots \ge ||f_{m+1}||$. If all $N(f_j), j = 1, \ldots, m+1$ are different, then we get by (1.1)

$$||f_1 + \cdots + f_{m+1}|| \le \max_{1 \le j \le m+1} (C_0^j ||f_j||)$$

  and

$$C_0^j ||f_j|| \le C_0 N(f_1) \le C_0(N(f_1)^p + \cdots + N(f_m)^p)^{1/p}.$$

  If $N(f_j) = N(f_{j+1}) = C_0^l$ for some $1 \le j \le m$ and some $l \in \mathbb{Z}$, then

$$||f_j + f_{j+1}|| \le C_0 \max(||f_j||, ||f_{j+1}||) \le C_0^{l+1}.$$

  Hence

$$N(f_j + f_{j+1})^p \le C_0^{(l+1)p} = 2^{l+1} = 2^l + 2^l = N(f_j)^p + N(f_{j+1})^p.$$

  Finally, we get by induction assumption

$$\begin{aligned} ||f_1 + \cdots + f_{m+1}|| &= ||f_1 + \cdots + f_{j-1} + (f_j + f_{j+1}) + f_{j+2} + \cdots + f_{m+1}|| \\ &\le C_0\Big(N(f_1)^p + \cdots + N(f_{j-1})^p + N(f_j + f_{j+1})^p + N(f_{j+2})^p + \cdots + N(f_{m+1})^p\Big)^{1/p} \\ &\le C_0\Big(N(f_1)^p + \cdots + N(f_{m+1})^p\Big)^{1/p}. \end{aligned}$$

□

1.3 Compactness and entropy numbers

Definition 6. Let X be a Banach space, or a p-Banach space or a quasi-Banach space.

a) $M \subset X$ is *open*, iff $\forall x \in M\ \exists \varepsilon > 0 :\ B(x,\varepsilon) = \{y \in X : ||x-y||_X < \varepsilon\} \subset M$.

b) $K \subset X$ is *compact*, iff $\forall \{M_\alpha\}_{\alpha\in I}, M_\alpha$ open with $\bigcup_{\alpha\in I} M_\alpha \supset K$ there exists a finite subsystem $\{a_1,\dots,\alpha_n\} \subset I$ with $\bigcup_{i=1}^{n} M_{\alpha_i} \supset K$.

c) $T \in \mathcal{L}(X,Y)$ is *compact*, iff $\overline{TB_X} \subset Y$ is compact.

Let $K \subset X$ be compact. Then[4] for every $\varepsilon > 0$, there are finitely many points $x_1,\dots,x_n$, such that

$$\bigcup_{i=1}^{n} B(x_i,\varepsilon) \supset K.$$

Of course, the number of points $n(\varepsilon)$ grows, as $\varepsilon \to 0$. The concept of dyadic entropy numbers works with the inverse function, i.e. we ask, how large balls do we need to take to cover the set K with only n of them.

Definition 7. Let X, Y be two quasi-Banach spaces and let $T \in \mathcal{L}(X,Y)$. The sequence

$$e_n(T) = \inf\{\varepsilon > 0 : \exists y_1,\dots,y_{2^{n-1}} : T(B_X) \subset \bigcup_{i=1}^{2^{n-1}} B(y_i,\varepsilon)\}, \qquad n \in \mathbb{N},$$

is called the sequence of *entropy numbers* of the operator T.

Remark 6. We shall sometimes write $y_i + \varepsilon B_Y$ or $B_Y(y_i,\varepsilon)$ instead of $B(y_i,\varepsilon)$.

The following theorem summarises the basic properties of entropy numbers.

Theorem 8. *Let X, Y, Z be three quasi-Banach spaces, let $S, T \in \mathcal{L}(X,Y)$ and let $R \in \mathcal{L}(Y,Z)$. Then it holds:*

(i) $||T|\mathcal{L}(X,Y)|| \ge e_1(T) \ge e_2(T) \ge \dots \ge 0$.
(ii) $e_n(T) \to 0$, *iff T is compact.*
(iii) $||T|\mathcal{L}(X,Y)|| = e_1(T)$ *if Y is a Banach space*[5]
(iv) $\forall n_1, n_2 \in \mathbb{N}$ *holds* $e_{n_1+n_2-1}(R \circ S) \le e_{n_1}(R) e_{n_2}(S)$.
(v) If Y is a p-Banach space, then

$$e^p_{n_1+n_2-1}(S+T) \le e^p_{n_1}(S) + e^p_{n_2}(T).$$

Proof. (i) The inequality $e_j(T) \ge e_{j+1}(T)$ follows directly from the Definition 7. We write $||T||$ instead of $||T|\mathcal{L}(X,Y)||$ for short. The inequality $||T|| \ge e_1(T)$ follows from

$$T(B_X) \subset B(0, ||T||) \subset Y.$$

[4] cf. Exercise 7a)
[5] cf. Exercise 10b.

(iii) If Y is a Banach space and $e_1(T) < ||T||$, then there exist some real number $0 < a < ||T||$ and some $y \in Y$, such that

$$T(B_X) \subset B(y,a),$$

i.e.

$$||Tx - y||_Y \le a$$

for all $x \in B_X$. Then for every $x \in B_X$ we obtain

$$||Tx||_Y \le \Big|\Big|\frac{Tx}{2} + \frac{y}{2}\Big|\Big|_Y + \Big|\Big|\frac{Tx}{2} - \frac{y}{2}\Big|\Big|_Y = \frac{1}{2}||T(-x) - y||_Y + \frac{1}{2}||Tx - y||_Y \le a,$$

hence $||T|| \le a$, which is a contradiction.

(ii) follows from Exercise 10a.

(iv) Let

$$R(B_Y) \subset \bigcup_{j=1}^{2^{n_1-1}} z_j + (e_{n_1}(R) + \varepsilon)B_Z,$$
$$S(B_X) \subset \bigcup_{i=1}^{2^{n_2-1}} y_i + (e_{n_2}(S) + \varepsilon)B_Y.$$

Then

$$\begin{aligned}(R \circ S)B_X = R(S(B_X)) \subset R\Bigg(\bigcup_{i=1}^{2^{n_2-1}} y_i + (e_{n_2}(S) + \varepsilon)B_Y\Bigg) &= \bigcup_{i=1}^{2^{n_2-1}} R(y_i) + (e_{n_2}(S) + \varepsilon)R(B_Y)\\ &\subset \bigcup_{i=1}^{2^{n_2-1}} \bigcup_{j=1}^{2^{n_1-1}} \underbrace{R(y_i) + (e_{n_2}(S) + \varepsilon)z_j}_{v_{i,j}} + (e_{n_1}(T) + \varepsilon)(e_{n_2}(S) + \varepsilon)B_Z\\ &= \bigcup_{i=1}^{2^{n_2-1}} \bigcup_{j=1}^{2^{n_1-1}} v_{i,j} + (e_{n_1}(T) + \varepsilon)(e_{n_2}(S) + \varepsilon)B_Z.\end{aligned}$$

Altogether, $(R \circ S)(B_X)$ may be covered by $2^{n_1-1+n_2-1} = 2^{(n_1+n_2-1)-1}$ balls in Z with radius $(e_{n_1}(T) + \varepsilon)(e_{n_2}(S) + \varepsilon)$. Finally, we let $\varepsilon \to 0$.

(v) follows similarly. Let

$$S(B_X) \subset \bigcup_{i=1}^{2^{n_1-1}} y_i + (e_{n_1}(S) + \varepsilon)B_Y,$$
$$T(B_X) \subset \bigcup_{j=1}^{2^{n_2-1}} z_j + (e_{n_2}(T) + \varepsilon)B_Y.$$

Then[6]

$$(S+T)(B_X) \subset S(B_X) + T(B_X) \subset \left(\bigcup_{i=1}^{2^{n_1-1}} y_i + (e_{n_1}(S)+\varepsilon)B_Y\right) + \left(\bigcup_{j=1}^{2^{n_2-1}} z_j + (e_{n_2}(T)+\varepsilon)B_Y\right)$$

$$\subset \bigcup_{i=1}^{2^{n_1-1}} \bigcup_{j=1}^{2^{n_2-1}} (y_i+z_j) + \left[(e_{n_1}(S)+\varepsilon)B_Y + (e_{n_2}(T)+\varepsilon)B_Y\right]$$

$$\subset \bigcup_{i=1}^{2^{n_1-1}} \bigcup_{j=1}^{2^{n_2-1}} (y_i+z_j) + \left[(e_{n_1}(S)+\varepsilon)^p + (e_{n_2}(T)+\varepsilon)^p\right]^{1/p} B_Y.$$

We have used Exercise 9 in the last step. Finally, we let $\varepsilon \to 0$ and observe, that

$$e_n(S+T) \le [e_{n_1}(S)^p + e_{n_2}(T)^p]^{1/p}$$

for n with $n-1 = n_1 - 1 + n_2 - 1$. □

1.4 Entropy numbers of $id : \ell_p^m \to \ell_q^m$

Up to very special cases (which we shall investigate in detail later on), the exact calculation of entropy numbers is almost impossible. Hence, we shall deal with estimates from above and below, which differ only through some constants, i.e. in formulas of the type

$$e_n(T) \approx n^{-1}, \qquad n \in \mathbb{N},$$

which means, that there are two positive constants c_1 and c_2, such that

$$c_1\, n^{-1} \le e_n(T) \le c_2 n^{-1}, \qquad n \in \mathbb{N}.$$

The most simple case, which demonstrate many of the significant properties of entropy numbers, is the operator

$$id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R}),$$

where $0 < p, q \le \infty$ and $m \in \mathbb{N}$. We are interested in estimates of $e_n(T)$ in the sense described above, but with c_1 and c_2 independent of n and m (but possibly depending on p and q).

Let us mention[7], that

$$e_n(id : \ell_p^m(\mathbb{C}) \to \ell_q^m(\mathbb{C})) \approx e_n(id : \ell_p^{2m}(\mathbb{R}) \to \ell_q^{2m}(\mathbb{R})), \qquad n, m \in \mathbb{N},$$

with constants of equivalence independent of n and m, but possibly depending on p and q. This somehow justifies our interest in real vector spaces.

We start with simple cases, it means with

Example 2. $id : \ell_\infty^m \to \ell_\infty^m$.

Step 1. Estimate from above.

We denote by $B = B_{\ell_\infty^m(\mathbb{R})} = [-1,1]^m$ the unit ball of $\ell_\infty^m(\mathbb{R})$.

[6] We denote by $A + B = \{a + b : a \in A, b \in B\}$
[7] cf. Exercise 16

We consider the sets

$$A_{k,m} = \left\{-\frac{2^k-1}{2^k}, -\frac{2^k-3}{2^k}, \dots, -\frac{1}{2^k}, \frac{1}{2^k}, \dots, \frac{2^k-1}{2^k}\right\}^m, \qquad k,m \in \mathbb{N}.$$

Simple calculation shows, that $A_{k,m}$ has

$$\left(\frac{2^k-1+(2^k-1)+2}{2}\right)^m = 2^{k\cdot m}$$

elements. Furthermore,

$$[-1,1] \subset \bigcup_{x\in A_{k,1}} x + \frac{1}{2^k}[-1,1],$$

and generally

$$B = \bigcup_{x\in A_{k,m}} x + \frac{1}{2^k}B.$$

This implies, that

$$e_{km+1}(id : \ell_\infty^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \le \frac{1}{2^k}.$$

So, if $n \in \mathbb{N}$ may be written as $n = km+1$, then

$$e_n(id) \le \frac{1}{2^k} = \frac{1}{2^{\frac{n-1}{m}}} = \frac{2^{1/m}}{2^{n/m}} \le 2 \cdot 2^{-n/m}.$$

The same estimate follows for all n by standard arguments, namely monotonicity. Let us describe this in detail. Let

$$\underbrace{k_0m+1}_{n_0} < n < \underbrace{(k_0+1)m+1}_{n_1=n_0+m} \tag{1.4}$$

for some $k_0 \ge 1$. Then

$$e_n(id) \le e_{n_0}(id) \le 2\cdot 2^{-n_0/m} = 2\cdot 2^{-\frac{n_1-m}{m}} \le 2\cdot 2^{-\frac{n-m}{m}} = 4\cdot 2^{-n/m}.$$

Finally, we consider $1 \le n \le m$, which cannot be expressed in the form given by (1.4). But for these n's we obtain trivially

$$e_n(id) \le 1 \le 2\cdot 2^{-n/m} \le 4\cdot 2^{-n/m}, \qquad 1 \le n \le m.$$

Step 2. Estimate from below.

We use volume arguments, which shall be very useful also later on. Let us assume, that

$$B \subset \bigcup_{j=1}^{2^{n-1}} y_j + \varepsilon B.$$

It means that

$$\text{vol } B = 2^m \le 2^{n-1}\varepsilon^m \cdot \text{vol } B = 2^{m+n-1}\varepsilon^m.$$

Hence

$$\varepsilon \ge 2^{\frac{1-n}{m}} \ge 2^{-n/m}.$$

Hence, we got

$$2^{-n/m} \le e_n(id : \ell_\infty^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \le 4\cdot 2^{-n/m}, \qquad n,m \in \mathbb{N}. \tag{1.5}$$

Let us mention, that even in this case, we got only estimates from above and from below, which differ by the constant 4.

The second simple case, namely $e_n(id : \ell_1^m(\mathbb{R}) \to \ell_1^m(\mathbb{R}))$ is postponed to the Exercise 11.

To be able to apply the volume arguments also in other (less trivial) situations, we shall need the Gamma function[8]

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} ds, \quad s > 0.$$

Theorem 9. *Let* $0 < p, q \le \infty$.

a) Then

$$\mathrm{vol}\ B_{\ell_p^m(\mathbb{R})} = 2^m \cdot \frac{\Gamma\big(1/p+1\big)^m}{\Gamma\big(m/p+1\big)}, \qquad m \in \mathbb{N}.$$

b) Then

$$e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) \ge 2^{\frac{1-n}{m}} \cdot \frac{\Gamma\big(1/p+1\big)}{\Gamma\big(1/q+1\big)} \cdot \left[\frac{\Gamma\big(m/q+1\big)}{\Gamma\big(m/p+1\big)}\right]^{1/m} \approx 2^{-n/m} \left[\frac{\Gamma\big(m/q+1\big)}{\Gamma\big(m/p+1\big)}\right]^{1/m}$$

with constants of equivalence independent of m and n, but depending on p and q.

c) Then

$$e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) \gtrsim 2^{-n/m} m^{1/q-1/p}.$$

Proof. Step 1. Proof of a)

We have

$$\mathrm{vol}\ B_{\ell_p^m(\mathbb{R})} = 2^m\ \mathrm{vol}\ (B_{\ell_p^m(\mathbb{R})} \cap [0,\infty)^m) = 2^m \int 1\ dx,$$

where the last integral goes over all $x = (x_1, \dots, x_m)$ with $x_1 \ge 0, \dots, x_m \ge 0$ and $x_1^p + \dots + x_m^p \le 1$. Through the substitution $t_j = x_j^p, \quad dt_j = p x_j^{p-1} dx_j$, this is equal to

$$\left(\frac{2}{p}\right)^m \int \prod_{j=1}^m t_j^{1/p-1}\ dt,$$

where the last integral goes over all $t = (t_1, \dots, t_m)$ with $t_1 \ge 0, \dots, t_m \ge 0$ and $t_1 + \dots + t_m \le 1$. This integral may be evaluated by induction, cf. Exercise 13, and this finishes the proof of a).

Step 2. Proof of b)

Let $B_{\ell_p^m(\mathbb{R})}$ be covered by 2^{n-1} ε-balls in $\ell_q^m(\mathbb{R})$ metric. Then

$$\mathrm{vol}\ B_{\ell_p^m(\mathbb{R})} \le 2^{n-1} \varepsilon^m \mathrm{vol}\ B_{\ell_q^m(\mathbb{R})}.$$

Hence,

$$\varepsilon \ge 2^{\frac{1-n}{m}} \left[\frac{\mathrm{vol}\ B_{\ell_p^m(\mathbb{R})}}{\mathrm{vol}\ B_{\ell_q^m(\mathbb{R})}}\right]^{1/m}.$$

This, together with a) gives b).

Step 3. Proof of c)

[8] cf. Exercise 12 for further information about Gamma (and Beta) function.

In view of b), it is enough to prove that

$$2^{-n/m}\left[\frac{\Gamma\big(m/q+1\big)}{\Gamma\big(m/p+1\big)}\right]^{1/m}\gtrsim 2^{-n/m}m^{1/q-1/p},$$

i.e.

$$\left[\frac{\Gamma\big(m/q+1\big)}{\Gamma\big(m/p+1\big)}\right]^{1/m}\gtrsim m^{1/q-1/p},\quad m\in\mathbb{N}. \tag{1.6}$$

We shall use the Stirling formula[9]

$$\Gamma(x)\approx\frac{1}{\sqrt{x}}\cdot\left(\frac{x}{e}\right)^x,\qquad x\geq 1.$$

Then

$$LHS\ (1.6)\approx\underbrace{\left(\frac{\frac{m}{p}+1}{\frac{m}{q}+1}\right)^{\frac{1}{2m}}}_{I}\cdot\underbrace{\left[\frac{\left(\frac{m}{q}+1\right)^{\frac{m}{q}+1}}{\left(\frac{m}{p}+1\right)^{\frac{m}{p}+1}}\right]^{1/m}}_{II}\cdot\underbrace{e^{-(\frac{m}{q}+1)\cdot\frac{1}{m}+(\frac{m}{p}+1)\cdot\frac{1}{m}}}_{III=e^{-1/q+1/p}\approx 1}.$$

First, we deal with I. Let us observe, that for $m=1$, I is equal to $\left[\frac{(1+p)q}{(1+q)p}\right]^{1/2}$ and for $m\to\infty$, I goes to 1. Hence, $I\approx 1$.

We rewrite II as

$$\begin{aligned}II&=\left(\frac{m}{q}\right)^{1/q+1/m}\cdot\left(\frac{m}{p}\right)^{-1/p-1/m}\cdot\underbrace{\left[\left(1+\frac{q}{m}\right)^{\frac{m}{q}+1}\right]^{1/m}\cdot\left[\left(1+\frac{p}{m}\right)^{\frac{m}{p}+1}\right]^{-1/m}}_{\approx 1,\text{again using limits}}\\&\approx m^{1/q-1/p}\cdot q^{-1/q}\cdot p^{1/p}\cdot(p/q)^{1/m}\approx m^{1/q-1/p}.\end{aligned}$$

□

1.4.1 Remark to interpolation theory

Let F be a linear operator with following properties:

$$F:L_2(\mathbb{R}^d)\to L_2(\mathbb{R}^d), \tag{1.7}$$

$$F:L_1(\mathbb{R}^d)\to L_\infty(\mathbb{R}^d). \tag{1.8}$$

Intuitively, it should follow that

$$F:L_p(\mathbb{R}^d)\to L_{p'}(\mathbb{R}^d),\qquad 1\leq p\leq 2,\qquad \frac{1}{p}+\frac{1}{p'}=1.$$

This is really the case and is the main subject of the so-called *interpolation theory* [10].

[9] The proof of this classical result goes beyond the scope of this script and we reffer to [3, Section 96, page 510] for details.

[10] In Jena, a lecture with exactly this title is sometimes offered by PD. D. D. Haroske.

Example 3. a) A prominent example of such an operator is given by the *Fourier transform*

$$\mathcal{F}(f)(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) e^{-ix\cdot\xi}\, dx.$$

Then (1.7) follows from the famous Parseval identity

$$||f|L_2(\mathbb{R}^d)|| = ||\mathcal{F}f|L_2(\mathbb{R}^d)||$$

and (1.8) follows almost trivially

$$|\mathcal{F}(f)(\xi)| \le \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} |f(x)| dx = \frac{||f|L_1(\mathbb{R}^d)||}{(2\pi)^{d/2}}.$$

b) Another famous application of the interpolation theory is the *convolution operator*.
Let $f \in L_1(\mathbb{R}^d)$ and put

$$M_f(g)(x) = \int_{\mathbb{R}^d} f(x-y) g(y) dy.$$

Then we get

$$||M_f(g)|L_1(\mathbb{R}^d)|| = \int_{\mathbb{R}^d} \big|M_f(g)(x)\big| dx \le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y)g(y)| dy dx = \int_{\mathbb{R}^d} |g(y)| \underbrace{\int_{\mathbb{R}^d} |f(x-y)| dx}_{\int_{\mathbb{R}^d} |f(x)| dx} dy$$
$$= \int_{\mathbb{R}^d} |f(x)| dx \cdot \int_{\mathbb{R}^d} |g(y)| dy = ||f|L_1(\mathbb{R}^d)|| \cdot ||g|L_1(\mathbb{R}^d)||.$$

And similarly,

$$\begin{aligned}||M_f(g)|L_\infty(\mathbb{R}^d)|| &\le \sup_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)g(y)| dy \le \sup_{x\in\mathbb{R}^d} \sup_{z\in\mathbb{R}^d} |g(z)| \int_{\mathbb{R}^d} |f(x-y)| dy \\ &= ||g|L_\infty(\mathbb{R}^d)|| \cdot ||f|L_1(\mathbb{R}^d)||.\end{aligned}$$

Hence

$$||M_f : L_1(\mathbb{R}^d) \to L_1(\mathbb{R}^d)|| \le ||f|L_1(\mathbb{R}^d)|| \quad \text{and} \quad ||M_f : L_\infty(\mathbb{R}^d) \to L_\infty(\mathbb{R}^d)|| \le ||f|L_1(\mathbb{R}^d)||.$$

By interpolation theory it follows, that

$$||M_f : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d)|| \le ||f|L_1(\mathbb{R}^d)||, \qquad 1 \le p \le \infty.$$

Theorem 10. *Let X be a quasi-Banach space, $\mathbb{Y}$ a vector space and let $T : X \to \mathbb{Y}$. Let $Y_0, Y_1 \subset \mathbb{Y}$ be two p-Banach spaces with $0 < p \le 1$, let $0 < \theta < 1$ and let $Y_\theta \subset \mathbb{Y}$ be a quasi-Banach space with $T \in \mathcal{L}(X, Y_0 \cap Y_1)$, where $||y|Y_0 \cap Y_1|| = \max(||y||_{Y_0}, ||y||_{Y_1})$. If*

$$||y|Y_\theta|| \le ||y|Y_0||^{1-\theta} \cdot ||y|Y_1||^\theta, \quad y \in Y_0 \cap Y_1 \subset Y_\theta, \tag{1.9}$$

then

$$e_{n_0+n_1-1}(T : X \to Y_\theta) \le 2^{1/p} e_{n_0}^{1-\theta}(T : X \to Y_0) \cdot e_{n_1}^\theta(T : X \to Y_1), \quad n_0, n_1 \in \mathbb{N}.$$

Proof. Let $a = (1+\varepsilon)e_{n_0}(T : X \to Y_0)$ and $b = (1+\varepsilon)e_{n_1}(T : X \to Y_1)$. It means, that there are $y_1, \ldots, y_{2^{n_0-1}} \in Y_0$ such that for every $x \in B_X$ there is a $j \in \{1, \ldots, 2^{n_0-1}\}$ with $||Tx - y_j||_{Y_0} \le a$. We consider the sets

$$B_j = \{x \in B_X : ||Tx - y_j||_{Y_0} \le a\} \subset B_X, \quad j = 1, \ldots, 2^{n_0-1}.$$

Obviously,

$$B_X = \bigcup_{j=1}^{2^{n_0-1}} B_j.$$

We can map each set B_j by T and cover in Y_1. So, there are points $z_j^1, \ldots, z_j^{2^{n_1-1}} \in Y_1$ such that for every $x \in B_j$ there is $i \in \{1, \ldots, 2^{n_1-1}\}$ such that $||Tx - z_j^i||_{Y_1} \le b$.

If z_j^i would lie in $T(B_j) \subset Y_0 \cap Y_1 \subset Y_\theta$, then we could use the calculation

$$||Tx - z_j^i||_{Y_\theta}^p \le \underbrace{||Tx - z_j^i||_{Y_0}^{p(1-\theta)}}_{\le(||Tx-y_j||_{Y_0}^p + ||y_j - z_j^i||_{Y_0}^p)^{1-\theta} \le (a^p + ||y_j - T(T^{-1}z_j^i)||_{Y_0}^p)^{1-\theta} \le (2a^p)^{1-\theta}} \cdot \underbrace{||Tx - z_j^i||_{Y_1}^{p\theta}}_{\le b^{p\theta}}.$$

Unfortunately, this is not necessarily the case. Hence, we choose

$$w_j^i \in B_{Y_1}(z_j^i, b) \cap T(B_j).$$

We may assume, that this is always possible. If not, then we just leave out z_j^i, because $B_{Y_1}(z_j^i, b)$ does not help with covering $T(B_j)$.

So, to every $x \in B_X$, we find $j \in \{1, \ldots, 2^{n_0-1}\}$ with $x \in B_j$ and then $i \in \{1, \ldots, 2^{n_1-1}\}$ such that

$$||Tx - w_j^i||_{Y_1}^p \le ||Tx - z_j^i||_{Y_1}^p + ||z_j^i - w_j^i||_{Y_1}^p \le 2b^p,$$

but also

$$||Tx - w_j^i||_{Y_0}^p \le ||Tx - y_j||_{Y_0}^p + ||y_j - w_j^i||_{Y_0}^p \le 2a^p,$$

hence

$$||Tx - w_j^i||_{Y_\theta} \le ||Tx - w_j^i||_{Y_0}^{1-\theta} \cdot ||Tx - w_j^i||_{Y_1}^{\theta} \le 2^{1/p} a^{1-\theta} b^\theta.$$

Finally, we let $\varepsilon \to 0$. □

Theorem 11. *Let $\mathbb{X}$ be a vector space and let $X_0, X_\theta, X_1 \subset \mathbb{X}$ with $0 < \theta < 1$ and $X_\theta \subset X_0 + X_1$.* [11] *We define the so-called Peetre interpolation K-functional by*

$$K(t, x) = \inf\{||x_0||_{X_0} + t||x_1||_{X_1} : x = x_0 + x_1, x_0 \in X_0, x_1 \in X_1\}, \qquad x \in X_\theta, \quad t > 0.$$

Let $T : D(T) \subset \mathbb{X} \to Y$, where Y is a p-Banach space, and let $T \in \mathcal{L}(X_0, Y)$ and $T \in \mathcal{L}(X_1, Y)$. If

$$t^{-\theta} K(t, x) \le ||x||_{X_\theta}, \quad t > 0, \quad x \in X_\theta, \tag{1.10}$$

then also $T \in \mathcal{L}(X_\theta, Y)$ and

$$e_{n_0+n_1-1}(T : X_\theta \to Y) \le 2^{1/p} e_{n_0}^{1-\theta}(T : X_0 \to Y) \cdot e_{n_1}^{\theta}(T : X_1 \to Y), \quad n_0, n_1 \in \mathbb{N}.$$

[11] This means, that every element $x \in X_\theta$ may be written as $x = x_0 + x_1$, where $x_0 \in X_0$ and $x_1 \in X_1$.

Proof. The boundedness of T as an operator from X_θ into Y follows easily. Let us take $x \in B_{X_\theta}$ and $\varepsilon > 0$. Then there are $x_0 \in X_0$ and $x_1 \in X_1$ such that $x = x_0 + x_1$ and $||x_0||_{X_0} + ||x_1||_{X_1} \le (1+\varepsilon)$ - just choose $t = 1$ in (1.10). The we have

$$||Tx||_Y^p \le ||Tx_0||^p + ||Tx_1||^p \le (1+\varepsilon)\left[||T|\mathcal{L}(X_0,Y)||^p + ||T|\mathcal{L}(X_1,Y)||^p\right],$$

hence (if we let $\varepsilon \to 0$)

$$||T|\mathcal{L}(X_\theta,Y)|| \le \left[||T|\mathcal{L}(X_0,Y)||^p + ||T|\mathcal{L}(X_1,Y)||^p\right]^{1/p}.$$

Let

$$a = e_{n_0}(T : X_0 \to Y), \quad b = e_{n_1}(T : X_1 \to Y), \quad t = b/a.$$

Let $x \in B_{X_\theta}$. Then there are $x_0 \in X_0$ and $x_1 \in X_1$ such that

$$||x_0||_{X_0} + t \cdot ||x_1||_{X_1} \le (1+\varepsilon)K(t,x) \le (1+\varepsilon)t^\theta ||x||_{X_\theta} \le (1+\varepsilon)t^\theta.$$

Let $y_1, \dots, y_{2^{n_0-1}} \in Y$ be a $(1+\varepsilon)a$ net for $T(B_{X_0})$ in Y and let $z_1, \dots, z_{2^{n_1-1}} \in Y$ be a $(1+\varepsilon)b$ net for $T(B_{X_1})$ in Y. Hence, there are j and k, such that

$$\left|\left|\frac{Tx_0}{(1+\varepsilon)t^\theta} - y_j\right|\right|_Y \le (1+\varepsilon)a \quad \text{and} \quad \left|\left|\frac{Tx_1}{(1+\varepsilon)t^{\theta-1}} - z_k\right|\right|_Y \le (1+\varepsilon)b.$$

We estimate

$$\begin{aligned} ||Tx - (1+\varepsilon)t^\theta y_j - (1+\varepsilon)t^{\theta-1}z_k||_Y^p &\le ||Tx_0 - (1+\varepsilon)t^\theta y_j||_Y^p + ||Tx_1 - (1+\varepsilon)t^{\theta-1}z_k||_Y^p \\ &\le [(1+\varepsilon)^2 a t^\theta]^p + [(1+\varepsilon)^2 b t^{\theta-1}]^p = 2(1+\varepsilon)^{2p}[a^{1-\theta}b^\theta]^p, \end{aligned}$$

i.e.

$$||Tx - (1+\varepsilon)t^\theta y_j - (1+\varepsilon)t^{\theta-1}z_k||_Y \le 2^{1/p}(1+\varepsilon)^2 a^{1-\theta}b^\theta.$$

We observe, that the set

$$\left\{(1+\varepsilon)t^\theta y_j + (1+\varepsilon)t^{\theta-1}z_k\right\}_{j,k}$$

forms a $2^{1/p}(1+\varepsilon)^2 a^{1-\theta}b^\theta$ net for $T(B_{X_\theta})$ in Y with cardinality $2^{n_0-1+n_1-1}$. □

Remark 7.
- In a typical situation, $\mathbb{Y}$ in Theorem 10 and $\mathbb{X}$ in Theorem 11 may be taken to be the space of all sequences or all measurable functions, respectively.
- Theorem 10 deals with interpolation on the target space. A following diagram is somtimes useful.

$$\begin{array}{ccc} & & Y_0 \\ & \nearrow & \\ T : X & \to & Y_\theta \\ & \searrow & \\ & & Y_1 \end{array}$$

The assumption (1.9) follows usualy with the help of Hölder's inequality, cf. Exercise 17.

- Theorem 11 deals with interpolation on the source space, this time with following diagram.

$$\begin{array}{ccc} T : X_0 & & \\ & \searrow & \\ T : X_\theta & \to & Y \\ & \nearrow & \\ T : X_1 & & \end{array}$$

  The assumption (1.10) is usualy more difficult to verify, cf. Exercise 18.

- Although the interpolation theory usually deals with interpolation on both (i.e. source and target) space side simultaneuously, there is no[12] analog of Theorem 10 and Theorem 11 for this situation. That would correspond to the diagram

$$\begin{array}{ccc} T : X_0 & \to & Y_0 \\ T : X_\theta & \to & Y_\theta \\ T : X_1 & \to & Y_1. \end{array}$$

1.4.2 Entropy numbers of $id : \ell_1^m \to \ell_\infty^m$

The main purpose of this section is to prove following

Theorem 12. *Let $m, n \in \mathbb{N}$. Then*

$$e_n(id : \ell_1^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \approx \begin{cases} 1, & if \quad 1 \le n \le \log_2 m, \\ \frac{\log\left(\frac{m}{n}+1\right)}{n}, & if \quad \log_2 m \le n \le m, \\ 2^{-n/m} m^{-1}, & if \quad m \le n, \end{cases}$$

where the constants of equivalence do not depend on m or n.

Proof. Step 1. $1 \le n \le \log_2 m$.

The estimate from above is trivial and follows from Theorem 8 (or just the fact, that $B_{\ell_1^m} \subset B_{\ell_\infty^m}$.)

Also the estimate from below is simple. Let us consider the canonical unit vectors

$$e^j = (e^j_1, \dots, e^j_m), \quad j = 1, \dots, m,$$

where

$$e^j_i = \begin{cases} 1, & \text{if} \quad i = j, \\ 0, & \text{otherwise}. \end{cases}$$

As $||e^j - e^k||_\infty = 1$ if $j \ne k$, each ℓ_∞^m-ball with radius strictly smaller than 1/2 contains (at most) one of these points. Hence, if $n \le \log_2 m$ (i.e. $2^n \le m$), then we need ℓ_∞^m balls of radius at least $1/2 \approx 1$ to cover $B_{\ell_1^m}$.

Step 2. $\log_2 m \le n \le m$.

[12] To be exact, there is some partial progress in this area connected mainly with the names of M. Cwikel, F. Cobos, T. Kühn and others, but the full solution is still missing.

Let

$$A_{k,m} = \{(t_1,\dots,t_m) \in \mathbb{Z}^m : ||t||_1 \le k\}, \qquad k,m \in \mathbb{N}.$$

The number of elements of $A_{k,m}$ may be estimated from above by (cf. Exercises 14 c and 15)

$$\#A_{k,m} \le {}^{13}2^k \cdot \#\{(t_0,t_1,\dots,t_m) \in \mathbb{N}_0^{m+1}, t_0+\dots+t_m = k\} = {}^{14}2^k \binom{k+m}{k} \le (2e)^k \cdot \left(\frac{k+m}{k}\right)^k.$$

Furthermore,

$$B_{\ell_1^m} \subset \bigcup_{t \in A_{k,m}} \frac{t}{k} + \frac{1}{k} B_{\ell_\infty^m}.$$

To prove this inclusion, we consider to each $x \in B_{\ell_1^m}$ an element $\tilde{x} = (\tilde{x}_1,\dots,\tilde{x}_m)$ with

$$\tilde{x}_i = \operatorname{sgn}(x_i) \cdot \frac{1}{k} \cdot \lfloor k|x_i| \rfloor^{15}, \qquad i = 1,\dots,m.$$

Then $||x_i - \tilde{x}_i||_\infty \le 1/k$ and $k\tilde{x} \in A_{k,m}$.

So, if

$$2^{n-1} \ge (2e)^k \left(1 + \frac{m}{k}\right)^k \quad \Longrightarrow \quad e_n(id : \ell_1^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \le \frac{1}{k}. \tag{1.11}$$

According to Exercise 19 a) there exists a constant $c > 0$, such that if

$$k \le cn \left[\log_2\left(1 + \frac{m}{n}\right)\right]^{-1}, \tag{1.12}$$

then

$$\frac{n}{k} \ge \log_2\left(12\left(1+\frac{m}{k}\right)\right),$$

hence

$$2^{\frac{n}{k}} \ge 12\left(1 + \frac{m}{k}\right)$$

and this again implies that

$$2^{n/k - 1/k} \ge 2e\left(1 + \frac{m}{k}\right)$$

from which the assumption of (1.11) follows. Hence, the conclusion of (1.11) holds for all k with (1.12). Choosing the k as big as possible finishes the proof of the estimate from above.

To prove the estimate from below, we estimate the number of elements of $A_{k,m}$ from below by

$$\# A_{k,m} \ge \#\{(t_0,t_1,\dots,t_m) \in \mathbb{N}_0^{m+1}, t_0+\dots+t_m = k\} = \binom{k+m}{k} \ge \left(\frac{k+m}{k}\right)^k.$$

So,

$$2^{n-1} \le \left(1 + \frac{m}{k}\right)^k \quad \Longrightarrow e_n(id : \ell_1^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \ge \frac{1}{2k}. \tag{1.13}$$

From Exercise 19 b) we know, that there is a constant $c' > 0$, such that if

$$k \ge c'n \left[\log_2\left(1 + \frac{m}{n}\right)\right]^{-1}, \tag{1.14}$$

[13] if $t_0 + \dots + t_m = k$, then at most k of $t_j's$ are different from zero. The factor 2^k corresponds to all possible signs.

[14] Just consider $m + k$ points on the real line and all the ways, in which you can scratch m of them. Then t_0 is the number of points before the first hole, t_1 the number of points between the first and the second hole, aso.

[15] By $\lfloor a \rfloor$ we denote the integer part of a real number a, i.e. $\lfloor a \rfloor \le a < \lfloor a \rfloor + 1$.

then

$$n \le k \log_2\Big(1+\frac{m}{k}\Big),$$

which further implies

$$2^{n-1} \le 2^n \le \Big(1+\frac{m}{k}\Big)^k.$$

So, every k with (1.14) satisfies also the conclusion of (1.13). Taking the smallest k possible finishes the proof.

Step 3. $m \le n$.

The estimate from below follows by volume arguments given in Theorem 9. We give the proof for the estimate from above.

We use the estimate of the number of elements of $A_{k,m}$

$$\#\ A_{k,m} \le 2^m \#\{(t_0,t_1,\dots,t_m)\in \mathbb{N}_0^{m+1}, t_0+\dots+t_m = k\} \le 2^m \binom{k+m}{m} \le (2e)^m \Big(1+\frac{k}{m}\Big)^m.$$

Then we have

$$2^{n-1} \ge (2e)^m\Big(1+\frac{k}{m}\Big)^m \quad \Longrightarrow \quad e_n(id: \ell_1^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \le \frac{1}{k}. \tag{1.15}$$

Let $n \ge 5m$. Then $2^{n/m-4} \ge 2$ and $2^{n/m-4} - 1 \ge 2^{n/m-5}$. So, if

$$\frac{k}{m} \le 2^{n/m-5} \le 2^{n/m-4} - 1 \le 2^{n/m-1/m-3} - 1,$$

we get

$$\Big(1+\frac{k}{m}\Big)^m \le 2^{n-1}2^{-3m} \le (2e)^{-m}2^{n-1}$$

and the conclusion of (1.15) follows. If $m \le n \le 5m$, the estimate follows by monotonicity.

Step 4. We give an alternative proof of the estimate from above for $m \le n$.

Let $0 < r < 1$ and let K_r be a maximal set of points from $B_{\ell_1^m}$, such that the mutual ℓ_∞^m-distance of every two different points is greater then r. This means

- $K_r = \{y_1,\dots,y_N\} \subset B_{\ell_1^m}$,
- $||y_i - y_j||_\infty > r$ for every $i \ne j$,
- for all $y \in B_{\ell_1^m}$ there is an $i \in \{1,\dots,N\}$ with $||y-y_i||_\infty \le r$.

Let us observe, that if $z \in B_{\ell_\infty^m}$, then

$$||y_j + rz||_1 \le ||y_j||_1 + r||z||_1 \le ||y_j||_1 + rm||z||_\infty \le 1 + rm,$$

which means, that

$$\bigcup_{j=1}^N y_j + rB_{\ell_\infty^m} \subset (1+rm)B_{\ell_1^m}.$$

Furthermore, if $i \ne j$, then

$$y_i + \frac{r}{2}B_{\ell_\infty^m} \cap y_j + \frac{r}{2}B_{\ell_\infty^m} = \emptyset.$$

This follows by contradiction; if $y_i + z_i = y_j + z_j$, with z_i and z_j from $\frac{r}{2} B_{\ell_\infty^m}$, then $||y_i - y_j||_\infty = ||z_i - z_j||_\infty \le r$, which is a contradiction with properties of the set K_r.
Comparing the volumes, we obtain

$$N \cdot \left(\frac{r}{2}\right)^m \cdot \text{vol } B_{\ell_\infty^m} \le (1 + rm)^m \text{vol } B_{\ell_1^m},$$

hence

$$N \le \frac{2^m (1 + rm)^m}{r^m m!}. \tag{1.16}$$

This leads to implication

$$N \le 2^{n-1} \quad \Longrightarrow \quad e_n(id : \ell_1^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \le r. \tag{1.17}$$

According to (1.16), this is the case, if

$$2^{n/m - 1/m} \ge \left(\frac{1}{r} + m\right) \cdot \frac{1}{\sqrt[m]{m!}} \approx \left(\frac{1}{r} + m\right) \cdot \frac{1}{m}.$$

So, if $m \le n$, we may put $\frac{1}{r} \approx m \, 2^{n/m}$, which finishes the proof. □

1.4.3 Extension to arbitrary p and q

Also this section has only one main aim, namely

Theorem 13. *a) Let $m, n \in \mathbb{N}$ and $0 < q \le p \le \infty$. Then*

$$e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) \approx m^{1/q-1/p} \cdot 2^{-n/m} \approx \begin{cases} m^{1/q-1/p}, & \text{if} \quad n \le m, \\ m^{1/q-1/p} 2^{-n/m}, & \text{if} \quad m \le n, \end{cases}$$

where the constants of equivalence do not depend on m or n.
b) Let $m, n \in \mathbb{N}$ and $0 < p \le q \le \infty$. Then

$$e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) \approx \begin{cases} 1, & \text{if} \quad 1 \le n \le \log_2 m, \\ \left[\frac{\log\left(\frac{m}{n}+1\right)}{n}\right]^{1/p-1/q}, & \text{if} \quad \log_2 m \le n \le m, \\ m^{1/q-1/p} 2^{-n/m}, & \text{if} \quad m \le n, \end{cases}$$

where the constants of equivalence do not depend on m or n.

Proof. [16]
a) The estimate from below follows from Theorem 9. The estimate from above for $n \le m$ follows by

$$e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) \le ||id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})|| = m^{1/q-1/p}.$$

We give the estimate from above for $n \ge m$ and $p = q$. It copies the Step 4. of the proof of Theorem 12.

[16] Full proof may be comming in some appendix...

Let again $0 < r < 1$ and let $K_r = \{y_1, \dots, y_N\}$ be a maximal set from $B_{\ell_p^m}$ with mutual ℓ_p^m-distance of points greater then r. We obtain again

$$\bigcup_{j=1}^{N} y_j + rB_{\ell_p^m} \subset 2B_{\ell_p^m}$$

and

$$y_i + \frac{r}{2}B_{\ell_p^m} \cap y_j + \frac{r}{2}B_{\ell_p^m} = \emptyset, \quad i \neq j.$$

Comparison of volumes leads to

$$N \cdot \left(\frac{r}{2}\right)^m \text{vol } B_{\ell_p^m} \leq 2^m \text{vol } B_{\ell_p^m},$$

i.e. $Nr^m \leq 4^m$. Hence if

$$r \geq \frac{8}{2^{n/m}} \geq \frac{4}{2^{(n-1)/m}},$$

we get $2^{n-1} \geq (4/r)^m \geq N$ and $e_n(id : \ell_p^m \to \ell_p^m) \leq r$ and the result follows.

If $0 < q \leq p \leq \infty$, then

$$\begin{aligned} e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) &\leq e_n(id : \ell_p^m(\mathbb{R}) \to \ell_p^m(\mathbb{R})) \cdot \underbrace{e_1(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R}))}_{\leq ||id:\ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})|| = m^{1/q-1/p}} \\ &\lesssim 2^{-n/m} m^{1/q-1/p}. \end{aligned}$$

b) Let $0 < p \leq q \leq \infty$.

The estimate from above for $1 \leq n \leq \log_2 m$ follows again by $B_{\ell_p^m} \subset B_{\ell_q^m}$. The estimate from below follows by considering the canonical unit vectors $e^1, \dots, e^m$.

If $\log_2 m \leq n \leq m$, we give the proof only for the Banach space setting, i.e. $1 \leq p \leq q \leq \infty$.

The estimate from above follows by interpolation

$$\begin{aligned} e_n(id : \ell_p^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) &\leq e_n^{1-\theta}(id : \ell_1^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \cdot e_1^{\theta}(id : \ell_\infty^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \\ &\lesssim \left[\frac{\log\left(1 + \frac{m}{n}\right)}{n}\right]^{1-\theta} \cdot 1^\theta = \left[\frac{\log\left(1 + \frac{m}{n}\right)}{n}\right]^{1/p}, \end{aligned}$$

where

$$\frac{1}{p} = \frac{1-\theta}{1} + \frac{\theta}{\infty}.$$

Then we interpolate on the target side

$$\begin{aligned} e_n(id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{R})) &\leq e_1^{1-\theta}(id : \ell_p^m(\mathbb{R}) \to \ell_p^m(\mathbb{R})) \cdot e_n^{\theta}(id : \ell_p^m(\mathbb{R}) \to \ell_\infty^m(\mathbb{R})) \\ &\lesssim \left[\frac{\log_2\left(1 + \frac{m}{n}\right)}{n}\right]^{\theta/p} = \left[\frac{\log_2\left(1 + \frac{m}{n}\right)}{n}\right]^{1/p-1/q}, \end{aligned}$$

with
$$\frac{1}{q} = \frac{1-\theta}{p} + \frac{\theta}{\infty}.$$

We skip the proof of the estimate from below. It is based on combinatorial arguments, quite similar to the case $p = 1, q = \infty$.

If $m \leq n$, the estimate from below is supplied by Theorem 9. The estimate from above follows similarly to the part a). We refer its proof into the Exercise 20. □

1.5 Eigenvalues and Carl-Triebel inequality

We restrict ourselves to the Banach space setting in this section. The generalisation to the quasi-Banach spaces may be found in the book [1].

Let X be a complex Banach space and let $T \in \mathcal{K}(X, X) = \mathcal{K}(X)$. Then the *spectrum* of T is defined as

$$\sigma(T) = \{\lambda \in \mathbb{C} : (T - \lambda I) \quad \text{is not boundedly invertible}\}.$$

Here, I is the identity mapping $I : X \to X$ and we say, that $(T - \lambda I)$ is boundedly invertible if

- $(T - \lambda I)^{-1}$ exists, i.e. $(T - \lambda I)$ must be injective and surjective, and
- $(T - \lambda I)^{-1}$ is boudned, i.e. $(T - \lambda I)^{-1} \in \mathcal{L}(X, X)$.

If for some $\lambda \in \mathbb{C}$, there is a $0 \neq x \in X$, such that $Tx = \lambda x$, then $(T - \lambda I)$ is not injective and hence $\lambda \in \sigma(T)$. Such a λ is called *eigenvalue* and the corresponding x is called *eigenvector* . But in general, not all the elements of $\sigma(T)$ are eigenvalues, cf. Exercise 21.

We recall briefly the *Riesz-Schauder theory* of compact operators.

If $T \in \mathcal{K}(X)$, then

- $\sigma(T)$ is countable,
- for all $\varepsilon > 0$, there are only finitely many $\lambda \in \sigma(T)$ with $|\lambda| \geq \varepsilon$,
- $0 \in \sigma(T)$,
- if $\lambda \in \sigma(T) \setminus \{0\}$, then λ is an eigenvalue and
- it has *finite multiplicity.*

We discuss in a bigger detail the notion of *multiplicity* of an eigenvalue. The *geometrical multiplicity* is defined as
$$\dim \ker\, (T - \lambda I)$$
and denotes the dimension of the space of eigenvectors associated to λ. The so-called *algebraic multiplicity* is defined as
$$\dim \bigcup_{k=1}^{\infty} \ker\, (T - \lambda I)^k$$
and is always bigger than (or equal to) the geometrical multiplicity. We refer to Exercise 24b.

According to the Riesz-Schauder theory, we may assign to each $T \in K(X)$ a sequence of all its eigenvalues
$$|\lambda_1(T)| \geq |\lambda_2(T)| \geq \cdots \geq 0,$$

where each eigenvalue is repeated with its algebraic multiplicity. If T has only finitely many eigenvalues, fill the rest of the sequence with zeros.

Theorem 14. *Let X be a complex Banach space and let $T \in \mathcal{K}(X)$. We re-order the eigenvalues, as described above. Then*

$$|\lambda_n(T)| \le \Big(\prod_{j=1}^n |\lambda_j(T)|\Big)^{1/n} \le \inf_{k\in\mathbb{N}} 2^{\frac{k}{2n}} e_k(T) \le \sqrt{2} e_n(T). \tag{1.18}$$

Proof. We give the proof in the most significant case, when all the eigenvalues are simple. The full proof may be found for example in the book [2].

So, take $n \in \mathbb{N}$ and $|\lambda_1(T)| \ge |\lambda_2(T)| \ge \cdots \ge |\lambda_n(T)| \ge 0$. Then there are linearly independent $x_1, \dots, x_n \in X$ such that $Tx_j = \lambda x_j, j = 1, \dots, n$. We define $M = \mathrm{span}(x_1, \dots, x_n)$. Then dim $M = n$ and $T(M) = M$.

Let us take $x \in M$, i.e. $x = \sum_{j=1}^n \gamma_j x_j$ with $\gamma_j \in \mathbb{C}, j = 1, \dots, n$ and $Tx = \sum_{j=1}^n \gamma_j \lambda_j x_j$.

We define an operator $J : M \to \mathbb{C}^n$, which assigns to each $x \in M$ the coefficients $\gamma_1, \dots, \gamma_n$, i.e.

$$Jx = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} \quad \text{and} \quad J^{-1} : \mathbb{C}^n \to M, \quad J^{-1}\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} = \sum_{j=1}^n \gamma_j x_j.$$

Then $Tx = J^{-1} T_n J x$ for every $x \in M$, where

$$T_n = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Unfortunately, we wish to apply volume arguments and in this context, $\mathbb{R}^{2n}$ seems to be more suitable space then $\mathbb{C}^n$.

Hence, we define $\mathbb{J} : M \to \mathbb{R}^{2n}$ and $\mathbb{J}^{-1} : \mathbb{R}^{2n} \to M$ by

$$\mathbb{J}x = \begin{pmatrix} Re(\gamma_1) \\ Im(\gamma_1) \\ \vdots \\ Re(\gamma_n) \\ Im(\gamma_n) \end{pmatrix} \quad \text{and} \quad \mathbb{J}^{-1}\begin{pmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_n \\ \beta_n \end{pmatrix} = \sum_{j=1}^n (\alpha_j + i\beta_j) x_j,$$

where $Re(z)$ denotes the real part of a complex number $z \in \mathbb{C}$ and $Im(z)$ its imaginary part. Then $Tx = \mathbb{J}^{-1}\mathbb{T}_n\mathbb{J}x$ for all $x \in M$, where

$$\mathbb{T}_n = \begin{pmatrix} Re(\lambda_1) & -Im(\lambda_1) & 0 & 0 & \dots & 0 & 0 \\ Im(\lambda_1) & Re(\lambda_1) & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & Re(\lambda_2) & -Im(\lambda_2) & \dots & 0 & 0 \\ 0 & 0 & Im(\lambda_2) & Re(\lambda_2) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & Re(\lambda_n) & -Im(\lambda_n) \\ 0 & 0 & 0 & 0 & \dots & Im(\lambda_n) & Re(\lambda_n) \end{pmatrix},$$

while

$$
\begin{aligned}
Tx &= \sum_{j=1}^{n} (Re(\lambda_j) + i\, Im(\lambda_j))(Re(\gamma_j) + i\, Im(\lambda_j))x_j \\
&= \sum_{j=1}^{n} \big\{ Re(\lambda_j)Re(\gamma_j) - Im(\lambda_j)Im(\gamma_j) + i\,(Re(\lambda_j)Im(\lambda_j) + Im(\lambda_j)Re(\lambda_j)) \big\} x_j.
\end{aligned}
$$

We define a measure μ on M by

$$\mu(K) = \text{vol}\,\Big(\mathbb{J}(K)\Big)$$

for all $K \subset M$, for which $\mathbb{J}(K) \subset \mathbb{R}^{2n}$ is Lebesgue-measurable. [17]

With the help of the notation introduced so far, the proof becomes simple.

For every $K \subset M$, we get

$$\mathbb{J}T(K) = \mathbb{J}\mathbb{J}^{-1}\mathbb{T}_n\mathbb{J}(K) = \mathbb{T}_n(\mathbb{J}(K)),$$

hence

$$
\begin{aligned}
\mu(T(K)) &= \text{vol}\,(\mathbb{J}T(K)) = \text{vol}\,(\mathbb{T}_n(\mathbb{J}(K))) \\
&= \det\,(\mathbb{T}_n)\cdot\ \text{vol}\,(\mathbb{J}(K)) \\
&= \begin{vmatrix} Re(\lambda_1) & -Im(\lambda_1) \\ Im(\lambda_1) & Re(\lambda_1) \end{vmatrix} \cdot \begin{vmatrix} Re(\lambda_2) & -Im(\lambda_2) \\ Im(\lambda_2) & Re(\lambda_2) \end{vmatrix} \cdots \begin{vmatrix} Re(\lambda_n) & -Im(\lambda_n) \\ Im(\lambda_n) & Re(\lambda_n) \end{vmatrix} \cdot \mu(K) \\
&= \left(\prod_{j=1}^{n} |\lambda_j|^2 \right) \cdot \mu(K).
\end{aligned}
$$

Let us mention, that this formula (with a slightly more technical proof, which uses the Jordan canonical form of T) holds also for eigenvalues with higher multiplicity.

Let $k \in \mathbb{N}$ and let

$$T(B_X) \subset \bigcup_{j=1}^{2^{k-1}} y_j + \varepsilon B_X,$$

where $\varepsilon > e_k(T)$ is arbitrary. Then

$$T(B_X \cap M) \subset \bigcup_{j=1}^{2^{k-1}} (y_j + \varepsilon B_X) \cap M.$$

If $(y_j + \varepsilon B_X) \cap M = \emptyset$, then we may leave out this j. Otherwise, we find $z_j \in M$, such that $y_j + \varepsilon B_X \subset z_j + 2\varepsilon B_X$. Then

$$T(B_X \cap M) \subset \bigcup_{j=1}^{2^{k-1}} (z_j + 2\varepsilon B_X) \cap M = \bigcup_{j=1}^{2^{k-1}} z_j + 2\varepsilon(B_X \cap M).$$

Comparing the μ-volumes, we obtain

$$\left(\prod_{j=1}^{n} |\lambda_j|^2 \right) \cdot \mu(B_x \cap M) = \mu(T(B_X \cap M)) \le 2^{k-1}\mu(2\varepsilon(B_X \cap M)) = 2^{k-1} \cdot (2\varepsilon)^{2n} \cdot \mu(B_X \cap M).$$

[17] Of course, vol denotes the Lebesgue measure on $\mathbb{R}^{2n}$.

Dividing by $\mu(B_X \cap M)$ and taking the $2n$-root, we get

$$\sqrt[n]{\prod_{j=1}^{n} |\lambda_j|} = \sqrt[2n]{\prod_{j=1}^{n} |\lambda_j|^{2n}} \le 2^{\frac{k-1}{2n}} \cdot 2\varepsilon \le 2 \cdot 2^{\frac{k}{2n}} \cdot \varepsilon.$$

As we may take ε arbitrarily close to $e_k(T)$, we get

$$\sqrt[n]{\prod_{j=1}^{n} |\lambda_j|} \le 2 \cdot 2^{\frac{k}{2n}} \cdot e_k(T).$$

We use the so-called *Carl's trick* to improve the constant, namely we apply the obtained result to T^r and $k' = kr$ and take $r \to \infty$. This leads to

$$\sqrt[n]{\prod_{j=1}^{n} |\lambda_j(T)|^r} = \sqrt[n]{\prod_{j=1}^{n} |\lambda_j(T^r)|} \le 2 \cdot 2^{\frac{kr}{2n}} e_{rk}(T^r) \le 2 \cdot 2^{\frac{kr}{2n}} e_k^r(T).$$

Taking the $1/r$ power gives

$$\sqrt[n]{\prod_{j=1}^{n} |\lambda_j(T)|} \le 2^{1/r} \cdot 2^{\frac{k}{2n}} \cdot e_k(T)$$

and we let $r \to \infty$ to finish the proof. □

1.6 Applications of entropy numbers

In this section, we describe two of the possible applications of compactness and entropy numbers. This concerns the area of signal processing and the spectral theory of partial differential equations.

1.6.1 Applications to signal processing

Let $T : X \to Y$ be compact, let $n \in \mathbb{N}$ and let $\varepsilon > e_n(T : X \to Y)$. Then

$$T(B_X) \subset \bigcup_{j=1}^{2^{n-1}} y_j + \varepsilon B_Y.$$

This means, that for every $x \in B_X$, there is a $j \in \{1, \dots, 2^{n-1}\}$ with $||Tx - y_j||_Y \le \varepsilon$.

We may define a mapping

$$f : B_X \to \{1, \dots, 2^{n-1}\}, \qquad f(x) = j \quad \text{with} \quad ||Tx - y_j||_Y \le \varepsilon.$$

If there are more j's with this property, we choose one of them arbitrarily. This means, that

- $||x - y_{f(x)}||_Y \le \varepsilon$ for all $x \in X$,
- we need only $n - 1$ bits to specify $f(x)$.

This may be used in an obvious way. If A and B are two different people or places and we want to transfer the information about an element $x \in B_X$ and are ready to accept accuracy $\varepsilon > 0$, then we need to transfer only $n-1$ bits, where $e_n(T) \le \varepsilon$. The price for this is, that both A and B has to know in advance the "lexicon" $y_1, \dots, y_{2^{n-1}}$.

We demonstrate this idea in connections with image processing. In this case, $x = \big(x_{i,j}\big)_{i,j=1}^N \in \mathbb{R}^{N^2}$, where $N \approx 1000$. For example, $x_{i,j} \in [-1,1]$ may give the gray scala going from $x_{i,j} = -1$ (black) to $x_{i,j} = 1$ (white). To deal with colorful pictures, we need the same approach for all the three RGB-channels, i.e. $m = 3N^2$.

It is a central observation of image processing, that expressing the vector $x = \{x_{i,j}\}_{i,j=1}^N \in \ell_2^{N^2}$ in a different orthonormal system leads to an element $\{y_{i,j}\}_{i,j=1}^N$ with the same $\ell_2^{N^2}$ norm, but (usually) with essentially smaller $\ell_p^{N^2}$ norm, where $p \le 2$ (usually even $p << 1$). This means, that only few of the coefficients are large and many of them are very small, cf. Picture 1.

We return to this orthonormal system later on, but show immediately, what it means for further processing of the image. Using n bits of information, we may decode the element $y \in \ell_2^{N^2}$ with the $\ell_2^{N^2}$-error smaller or equal to

$$||y||_{\ell_p^{N^2}} \cdot e_n(id : \ell_p^{N^2}(\mathbb{R}) \to \ell_2^{N^2}(\mathbb{R})) \lesssim ||y||_{\ell_p^{N^2}} \cdot \left[\frac{\log_2\big(1 + \frac{N^2}{n}\big)}{n}\right]^{1/p-1/2},$$

where we have assumed that

$$\log_2 N^2 \le n \le N^2, \qquad \text{i.e.} \quad 2 \cdot \log_2 1024 = 20 \le n \le 1024^2,$$

which is definitely the typical case. We observe, that

- The success of the compression depends on the picture - "simple" picture is supposed to have $||y||_{\ell_p^{N^2}}$ small also for $p << 1$ small.
- The decay of the error is surprisingly fast and improves with p getting smaller, for example (up to the logarithmical factor) n^{-2} for $p = 2/5$.

1.6.2 Haar bases

We return to the construction of the orthonormal system, which (at least for "simple" pictures) should lead to smaller ℓ_p-quasi-norms.

First, we construct a Haar basis on $[-1,1]^d$. We start with $d = 1$ and set

$$h_0^0(t) = 1, \quad t \in [-1,1]$$

and

$$h_1^0(t) = \begin{cases} 0, & \text{if} \quad |t| > 1, \\ 1, & \text{if} \quad -1 \le t \le 0, \\ -1, & \text{if} \quad 0 < t \le 1. \end{cases}$$

The function h_0^0 is usually called *father wavelet* and the function h_1^0 is called *mother wavelet* . All the other vectors of the Haar basis are derived from the mother wavelet by

$$h_j^i(t) = h_1^0(2^{j-1}t + 2^{j-1} - (2i+1)), \qquad j \in \{2,3,\dots\}, \quad i \in \{0,1,\dots,2^{j-1}-1\}.$$

It is easy to see, that

$$\{h_j^i\} \quad \text{with} \quad j = i = 0 \quad \text{or} \quad j \in \{1, 2, 3, \dots\}, \quad i \in \{0, 1, \dots, 2^{j-1} - 1\}$$

are mutualy orthogonal. So, after a proper normalisation, we obtain an orthonormal basis in $L_2([-1, 1])$.

If $d > 1$, the situation becomes more interesting. There are namely two ways, how to generalise this construction to higher dimensions. The first is to consider the tensor products.

Let $\varphi^1(t), \dots, \varphi^d(t) \in L_2([-1, 1])$. Then we set

$$(\varphi^1 \otimes \dots \otimes \varphi^d)(t_1, \dots, t_d) = \varphi^1(t_1) \dots \varphi^d(t_d), \quad (t_1, \dots, t_d) \in [-1, 1]^d.$$

The following lemma then provides the way to construct an orthonormal basis in higher dimensions.

Lemma 15. *Let*

$$\{\varphi_j^i\}, \qquad j \in \mathbb{N}_0$$

be for each $i = 1, \dots, d$ an orthonormal basis of $L_2([-1, 1])$. Then

$$\{\varphi_{j_1}^1 \otimes \dots \otimes \varphi_{j_d}^d\}, \quad j = (j_1, \dots, j_d) \in \mathbb{N}_0^d$$

is an orthonormal basis of $L_2([-1, 1]^d)$.

We leave the proof of this lemma to the Exercises.

The second construction works with several mother wavelets. We present the main idea in $d = 2$. We set

$$h_0(t_1, t_2) = 1, \quad -1 \le t_1, t_2 \le 1$$

and

$$h_{1,1}(t_1, t_2) = \begin{cases} -1, & \text{if} \quad 0 \le t_1 \le 1, |t_2| \le 1, \\ 1, & \text{if} \quad -1 \le t_1 < 0, |t_2| \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

$$h_{1,2}(t_1, t_2) = \begin{cases} -1, & \text{if} \quad 0 \le t_2 \le 1, |t_1| \le 1, \\ 1, & \text{if} \quad -1 \le t_2 < 0, |t_1| \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

$$h_{1,3}(t_1, t_2) = \begin{cases} -1, & \text{if} \quad -1 \le t_1 < 0, -1 \le t_2 < 0, \\ -1, & \text{if} \quad 0 \le t_1 \le 1, 0 \le t_2 \le 1, \\ 1, & \text{if} \quad -1 \le t_1 < 0, 0 \le t_2 \le 1, \\ 1, & \text{if} \quad -1 \le t_2 < 0, 0 \le t_1 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

This definition is motivated by the observation, that the matrices

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

form an orthogonal basis of $\mathbb{R}^{2\times 2}$.

The other wavelets are again produced from the mother wavelets, namely by the formula

$$h_{j,k}^{i_1,i_2}(t_1,t_2) = h_{1,k}(2^{j-1}(t_1,t_2) + (2^{j-1} - (2i_1+1), 2^{j-1} - (2i_2+1))),$$

where

$$j \in \{2,3,\dots\}, \quad i_1, i_2 \in \{0,1,\dots,2^{j-1}-1\}, \quad k \in \{1,2,3\}.$$

We put also $h_{1,1}^{0,0} = h_{1,1}$, $h_{1,2}^{0,0} = h_{1,2}$, $h_{1,3}^{0,0} = h_{1,3}$. Now the orthogonal basis in $L_2([-1,1]^2)$ is

$$\{h_0\} \cup \{h_{j,k}^{i_1,i_2}\}_{i_1,i_2,j,k}, \quad \text{with} \quad j \in \{1,2,3,\dots\}, \quad i_1,i_2 \in \{0,1,\dots,2^{j-1}-1\}, \quad k \in \{1,2,3\}.$$

Both these constructions may be directly generalised to sequence spaces. For example, for $\mathbb{R}^{4\times 4}$ we consider the father-matrix and three mother-matrices

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

together with

$$\begin{pmatrix} -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

and further 2×4 matrices for the other two mother wavelets, i.e. alltogether 16 matrices.

Remark 8. Surprisingly enough, the second approach (based on several mother wavelets rather than on tensor constructions) is easier to handle and is usualy used in the literature.

1.6.3 Applications to PDE's

In this section, we present applications to (partial) differential equations, which are based on the Carl-Triebel inequality (1.18). First, we show on one example the importance of eigenvalues in mathematical physics.

Let

- $\Omega \subset \mathbb{R}^2$ be a C^∞ domain,
- $\Gamma = \partial\Omega$,
- $p(x)$ be the external force,
- $v \in C^2(\Omega)$ $\mathrm{tr}_\Gamma v = 0$ be the elongation of the membrane fixed on Γ and caused by the force $p(x)$,
- F be the surface described by $v(x)$, i.e. $F = \{(x_1,x_2,x_3) : v(x_1,x_2) = x_3\}$.

We denote by $\Delta F = |F| - |\Omega|$ the enlargement of the area caused by the force p. The resulting function v is given by an interplay between the force p and the inner force of the membrane (which depends on ΔF). This idea is mirrored in the corresponding potential

$$J(v) = \Delta F - \int_\Omega p(x)v(x)dx = \int_\Omega \underbrace{\sqrt{1 + \left(\frac{\partial v}{\partial x_1}(x)\right)^2 + \left(\frac{\partial v}{\partial x_2}(x)\right)^2}}_{\to |F|} - \underbrace{1}_{\to|\Omega|} - p(x)v(x)dx.$$

If both the partial derivatives are small, we may use the formula $\sqrt{1+x} \approx 1 + \frac{x}{2}$ and get

$$J(v) \approx \tilde{J}(v) = \int_\Omega \frac{1}{2}|\nabla v(x)|^2 - p(x)v(x)dx.$$

We are looking for a function v, where the potential $\tilde{J}(v)$ is minimal. If φ is an arbitrary smooth function (for example $\varphi \in C_0^2(\Omega)$ or $\varphi \in C_0^\infty(\Omega)$), then it must hold

$$\tilde{J}(v + \varepsilon\varphi) \geq \tilde{J}(v), \quad \varepsilon \in \mathbb{R}.$$

This leads to

$$\begin{aligned}
0 &= \frac{d}{d\varepsilon}\tilde{J}(v+\varepsilon\varphi)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon}\Big(\int_\Omega \frac{1}{2}|\nabla v(x) + \varepsilon\nabla\varphi(x)|^2 - p(x)(v(x)+\varepsilon\varphi(x))dx\Big) \\
&= \frac{d}{d\varepsilon}\Big(\int_\Omega \frac{1}{2}|\nabla v(x)|^2 - p(x)v(x)dx + \varepsilon\int_\Omega \big(\frac{\partial v}{\partial x_1}\cdot\frac{\partial \varphi}{\partial x_1} + \frac{\partial v}{\partial x_2}\cdot\frac{\partial \varphi}{\partial x_2}\big) - p(x)\varphi(x)dx \\
&\quad + \frac{\varepsilon^2}{2}\int_\Omega |\nabla\varphi(x)|^2 dx\Big)\Big|_{\varepsilon=0} \\
&= 0 + \int_\Omega (\nabla v)(x)\cdot(\nabla\varphi)(x) - p(x)\varphi(x)dx + 0
\end{aligned}$$

This means, that

$$\int_\Omega (\nabla v)(x)\cdot(\nabla\varphi)(x) - p(x)\varphi(x)dx = 0$$

for every φ from (let us say) $C_0^\infty(\Omega)$. Using the Green's theorem, this is equivalent to

$$\int_\Omega \varphi(x)(-\Delta v(x) - p(x))dx = 0$$

and as this is supposed to hold for every φ, we arrive to the *Dirichlet problem*

$$\begin{aligned}
\Delta v(x) = -p(x), \quad x \in \Omega, \\
v\big|_{\partial\Omega} = 0.
\end{aligned} \tag{1.19}$$

The solution of (1.19) is sometimes called *stationary solution.* It means, that if the membrane is in this stage and does not move (position is equal to v, velocity is equal to zero), then the combined action of the external force p and the action of the internal force caused by tension in F cancel each other. Or we may put it other way: if the surface F is described by v, then action of the tension of F is the same as the action of an external force $p(x)$ given by $p(x) = \Delta v(x)$.[18]

Now we consider, that the membrane oscilates (hence $v = v(x,t)$ depends also on the time $t > 0$) and there is no external force p. The tension of the surfaces gives the surface acceleration

$$m(x)\cdot\frac{\partial^2 v(x,t)}{\partial t^2},$$

where $m(x)$ is the mass density (which we shall put equal to 1 for simplicity).

This leads to the equation

$$\Delta_x v(x,t) = \frac{\partial^2 v(x,t)}{\partial t^2}.$$

[18] Note the *plus* sign!

We look first for solutions of the type

$$v(x,t) = e^{i\lambda t}u(x),$$

where

- λ is the main and only time frequency of this solution,
- $u(x)$ is its amplitude at the point $x \in \Omega$,
- $u(x) = 0$ for all $x \in \partial\Omega$.

This leads immediately to

$$e^{i\lambda t}\Delta u(x) = -\lambda^2 e^{i\lambda t}u(x),$$

i.e.

$$\Delta u(x) = -\lambda^2 u(x), \quad u\big|_{\partial\Omega} = 0. \tag{1.20}$$

The general solution is then a linear combination of these particular solutions $u_\lambda(x)e^{i\lambda x}$. The addmisible λ's (also those λ's for which a particular solution of (1.20) exists) are called *eigenfrequencies*.

Remark 9. For $\Omega = B(0,1)$ (and some other special domains) we may solve the problem explicitly. For general domains it is impossible. Nevertheless, one would still like to compute/estimate the eigenfrequencies.

When dealing with this question, we may immediately observe two problems.

First problem.

The operators

$$d_2 : L_2([-1,1]) \to L_2([-1,1]), \quad d_2(u) = u''$$

or

$$D_2 : L_2(\Omega) \to L_2(\Omega), \quad D_2(u) = \Delta u$$

are not even boudned (and hence very far from being compact).

A huge portion of functional analysis is therefore devoted to the study of *unbounded operators*. These are operators defined on a subspace of some Banach space X with values in the same space. For example

$$T : \operatorname{dom}(T) \subset L_2([-1,1]) \to L_2([-1,1]),$$

where

$$\operatorname{dom}(T) = \{f \in C^2([-1,1]) : f(-1) = f'(-1) = 0\}$$

and

$$Tf = f'' \in C([-1,1]) \subset L_2([-1,1]).$$

If $Tf(x) = f''(x) = g(x)$, we get $f'(x) = \int_{-1}^x g(t)dt$ and

$$f(x) = \int_{-1}^x \int_{-1}^u g(t)dtdu = \int_{-1}^x g(t)\int_t^x 1dudt = \int_{-1}^x g(t)\cdot(x-t)dt.$$

This means, that

$$(T^{-1}g)(x) = \int_{-1}^x g(t)\cdot(x-t)dt, \quad T^{-1} : L_2([-1,1]) \to L_2([-1,1])$$

is the inverse of T and the domain of T may be enlarged to

$$\mathrm{dom}'(T) := \{f \in L_2([-1,1]) : \exists g \in L_2([-1,1]) \quad \text{with} \quad T^{-1}g = f\}.$$

One observes that

- The eigenvalues λ of T are the reciprocal values of eigenvalues of T^{-1},
- T^{-1} is compact.

At this stage, one may apply Carl-Triebel inequality to T^{-1}.

Second problem.

The second problem is that we were (almost) able to calculate the entropy numbers of $id : \ell_p^m(\mathbb{R}) \to \ell_q^m(\mathbb{C})$, but to estimate the entropy numbers of T, we need some information about *function spaces*. Obviously, it would be very useful to find some way, how to transfer our results about sequence space to function spaces. The most simple example of such an approach are the *Fourier series* in $L_2([-\pi,\pi])$.

The set

$$B = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos nx, \frac{1}{\sqrt{\pi}} \sin nx \right\}_{n=1}^{\infty}$$

is an orthonormal basis of $L_2(-\pi,\pi)$. This means, that

$$(f,g) = \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx = 0$$

for all $f,g \in B, f \neq g$ and $(f,f) = 1$ for all $f \in B$.

Let f be a measurable function on $(-\pi,\pi)$. Then $f \in L_2(-\pi,\pi)$ if, and only if, there are two sequences $\{a_n\}_{n=0}^{\infty} \in \ell_2$ and $\{b_n\}_{n=0}^{\infty} \in \ell_2$, such that

$$f(x) = \frac{a_0}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} a_n \frac{\cos nx}{\sqrt{\pi}} + b_n \frac{\sin nx}{\sqrt{\pi}},$$

with convergence in $L_2(-\pi,\pi)$ and

$$||f|L_2(-\pi,\pi)|| = \left(\sum_{n=0}^{\infty} a_n^2 + \sum_{n=1}^{\infty} b_n^2 \right)^{1/2} = \left(||a|\ell_2||^2 + ||b|\ell_2||^2 \right)^{1/2}.$$

In the same way, one may consider the system

$$B^1 = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos nx}{\sqrt{(1+n^2)\pi}}, \frac{\sin nx}{\sqrt{(1+n^2)\pi}} \right\}_{n=1}^{\infty}.$$

This is an orthonormal base in the first order Sobolev space

$$W_2^1(-\pi,\pi) = \{f \in AC(-\pi,\pi) : f, f' \in L_2(-\pi,\pi)\}$$

equipped with the scalar product

$$(f,g)_{W_2^1} = \int_{-\pi}^{\pi} f(x)\overline{g(x)} + f'(x)\overline{g'(x)}dx.$$

Again, $f \in W_2^1(-\pi,\pi)$ if, and only if, there exist two sequences $\{\alpha_n\}_{n=0}^\infty \in \ell_2$ and $\{\beta_n\}_{n=0}^\infty$, such that

$$f(x) = \frac{\alpha_0}{\sqrt{2\pi}} + \sum_{n=1}^\infty \alpha_n \frac{\cos nx}{\sqrt{(1+n^2)\pi}} + \beta_n \frac{\sin nx}{\sqrt{(1+n^2)\pi}}$$

with convergence in $W_2^1(-\pi,\pi)$.

Using the fact, that B and B' differ only by constant, we get

$$a_n = \frac{\alpha_n}{\sqrt{1+n^2}}, \quad b_n = \frac{\beta_n}{\sqrt{1+n^2}}.$$

Hence,

$$||f|W_2^1(-\pi,\pi)|| = \left(\sum_{n=0}^\infty \alpha_n^2 + \sum_{n=1}^\infty \beta_n^2\right)^{1/2} = \left(\sum_{n=0}^\infty (1+n^2)a_n^2 + \sum_{n=1}^\infty (1+n^2)b_n^2\right)^{1/2}.$$

Remark 10.

- This may be generalised to define

$$||f|W_2^s(-\pi,\pi)|| := \left\{ f \in L_2(-\pi,\pi) : \left(a_0^2 + \sum_{n=1}^\infty (1+n^2)^s(a_n^2+b_n^2)\right)^{1/2} < \infty \right\}.$$

  for arbitrary $s \geq 0$.

- For $p \neq 2$ goes everything wrong. Especially $L_p \neq \{f : (\sum a_n^p + b_n^p)^{1/p} < \infty\}$.

- One may proceed very similar (using tensor products) for $W_2^s((-\pi,\pi)^d)$, but $W_2^s(\Omega)$ needs essentially new ideas.

- The embedding $id : W_2^1(-\pi,\pi) \to L_2(-\pi,\pi)$ is compact, cf. Exercise 26c.

We show, how we may use our results obtained so far, to estimate the entropy numbers of

$$e_n(id : W_2^1(-\pi,\pi) \to L_2(-\pi,\pi)), \qquad n \in \mathbb{N}.$$

We denote by

$$S^0 : L_2(-\pi,\pi) \to \ell_2(\mathbb{Z})$$

the operator, which assigns to each f its Fourier coefficients in the base B, i.e.

$$(S^0 f)(n) = \begin{cases} \int_{-\pi}^{\pi} \frac{1}{\sqrt{2\pi}} \cdot f(x)dx, & \text{for} \quad n = 0, \\ \int_{-\pi}^{\pi} \frac{\cos nx}{\sqrt{\pi}} f(x)dx, & \text{for} \quad n \in \mathbb{N}, \\ \int_{-\pi}^{\pi} \frac{\sin nx}{\sqrt{\pi}} f(x)dx, & \text{for} \quad -n \in \mathbb{N}, \end{cases}$$

and by

$$S^1 : W_2^1(-\pi,\pi) \to \ell_2(\mathbb{Z})$$

a similar operator with respect to B', i.e.

$$(S^1 f)(n) = \frac{(S^0 f)(n)}{\sqrt{1+n^2}} = \begin{cases} \displaystyle\int_{-\pi}^{\pi} \frac{1}{\sqrt{2\pi}} \cdot f(x)dx, & \text{for} \quad n = 0, \\ \displaystyle\int_{-\pi}^{\pi} \frac{\cos nx}{\sqrt{(1+n^2)\pi}} f(x)dx, & \text{for} \quad n \in \mathbb{N}, \\ \displaystyle\int_{-\pi}^{\pi} \frac{\sin nx}{\sqrt{(1+n^2)\pi}} f(x)dx, & \text{for} \quad -n \in \mathbb{N}. \end{cases}$$

Hence, we have a following commutative diagrams

$$\begin{array}{ccc} W_2^1(-\pi,\pi) & \xrightarrow{id} & L_2(-\pi,\pi) \\ S^1 \big\downarrow & & \big\uparrow (S^0)^{-1} \\ \ell_2(\mathbb{Z}) & \xrightarrow{D} & \ell_2(\mathbb{Z}), \end{array} \qquad \begin{array}{ccc} W_2^1(-\pi,\pi) & \xrightarrow{id} & L_2(-\pi,\pi) \\ (S^1)^{-1} \big\uparrow & & \big\downarrow S^0 \\ \ell_2(\mathbb{Z}) & \xrightarrow{D} & \ell_2(\mathbb{Z}), \end{array}$$

where the *diagonal operator* $D : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ is defined as

$$D(\{x_n\}_{n\in\mathbb{Z}}) = \left\{ \frac{x_n}{\sqrt{1+n^2}} \right\}_{n\in\mathbb{Z}}.$$

The first diagram leads to

$$\begin{aligned} e_n(id) &= e_n((S^0)^{-1} \circ D \circ S^1) \le ||(S^0)^{-1}|\mathcal{L}(\ell_2(\mathbb{Z}), L_2(-\pi,\pi))|| \cdot e_n(D) \cdot ||S^1|\mathcal{L}(W_2^1(-\pi,\pi), \ell_2(\mathbb{Z}))|| \\ &= e_n(D) \end{aligned}$$

and the second to

$$\begin{aligned} e_n(D) &= e_n(S^0 \circ id \circ (S^1)^{-1}) \le ||S^0|\mathcal{L}(L_2(-\pi,\pi), \ell_2(\mathbb{Z}))|| \cdot e_n(id) \cdot ||(S^1)^{-1}|\mathcal{L}(\ell_2(\mathbb{Z}), W_2^1(-\pi,\pi))|| \\ &= e_n(id), \end{aligned}$$

hence $e_n(D) = e_n(id)$.

We shall consider the (notationaly simpler) operator $D' : \ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})$ defined by

$$D'(\{x_n\}_{n\in\mathbb{N}}) = \left\{ \frac{x_n}{n} \right\}_{n\in\mathbb{N}}$$

and show, that $e_n(D') \approx \frac{1}{n}$. We split $D' = D_n + D^n$, where

$$D_n(\{x_n\}_{n\in\mathbb{N}}) = \left(\frac{x_1}{1}, \frac{x_2}{2}, \dots, \frac{x_n}{n}, 0, 0, \dots \right), \qquad D^n = D' - D_n.$$

The estimate from below follows very quickly

$$e_n(D') \ge e_n(D_n) \ge \frac{1}{n} \cdot e_n(id : \ell_2^n \to \ell_2^n) \approx \frac{1}{n},$$

where we used Theorem 13 in the last step.

For the estimate from above, we use first

$$e_n(D') \le e_n(D_n) + ||D^n|| \le e_n(D_n) + \frac{1}{n}.$$

To estimate $e_n(D_n)$ from above, we assume, that $n = 2^m, m \in \mathbb{N}$ and split D_n into dyadic blocks $D_n = \Delta_1 + \Delta_2 + \dots + \Delta_m$, where

$$\begin{aligned}
\Delta_1 &: \{x_n\}_{n\in\mathbb{N}} \to \left(\frac{x_1}{1}, \frac{x_2}{2}, 0, 0, \dots\right),\\
\Delta_2 &: \{x_n\}_{n\in\mathbb{N}} \to \left(0, 0, \frac{x_3}{3}, \frac{x_4}{4}, 0, 0, \dots\right),\\
\Delta_3 &: \{x_n\}_{n\in\mathbb{N}} \to \left(0, 0, 0, 0, \frac{x_5}{5}, \frac{x_6}{6}, \frac{x_7}{7}, \frac{x_8}{8}, 0, 0, \dots\right),\\
\Delta_4 &: \{x_n\}_{n\in\mathbb{N}} \to \left(0, \dots, 0, \frac{x_9}{9}, \dots, \frac{x_{16}}{16}, 0, 0, \dots\right),\\
&\vdots\\
\Delta_m &: \{x_n\}_{n\in\mathbb{N}} \to \left(0, \dots, 0, \frac{x_{2^{m-1}+1}}{2^{m-1}+1}, \dots, \frac{x_{2^m}}{2^m}, 0, 0, \dots\right),
\end{aligned}$$

and use the subaditivity of entropy numbers, i.e.

$$e_{cn}(D_n) \le e_{n_1}(\Delta_1) + e_{n_2}(\Delta_2) + \dots + e_{n_m}(\Delta_m).$$

We need to choose the natural numbers $n_1, \dots n_m$ so, that

$$\sum_{j=1}^m n_j \le c\; 2^m$$

and

$$\sum_{j=1}^m e_{n_j}(\Delta_j) \le c' \frac{1}{2^m}.$$

Let us choose $n_j = (1+\varepsilon)(m-j+1)2^{j-1}$, $j = 1, \dots, m$.[19] Then it holds ($\varepsilon > 0$ is fixed and does *not* tend to zero):

$$\sum_{j=1}^m n_j = (1+\varepsilon)\sum_{j=1}^m (m-j+1)2^{j-1} \underset{l=m-j+1}{=} (1+\varepsilon)\sum_{l=1}^m l2^{m-l+1} = (1+\varepsilon)2^{m+1}\sum_{l=1}^m l2^{-l} \le c(1+\varepsilon)2^m$$

and by Theorem 13

$$\begin{aligned}
\sum_{j=1}^m e_{n_j}(\Delta_j) &\le e_{n_1}(id : \ell_2^2 \to \ell_2^2) + \sum_{j=2}^m \frac{e_{n_j}(id : \ell_2^{2^{j-1}} \to \ell_2^{2^{j-1}})}{2^{j-1}}\\
&\le c \sum_{j=1}^m 2^{-n_j/2^{j-1}} \cdot \frac{1}{2^{j-1}} = c \sum_{j=1}^m 2^{-(1+\varepsilon)(m-j)-j+1}\\
&= c'\, 2^{-(1+\varepsilon)m} \sum_{j=1}^m 2^{j\varepsilon} \approx 2^{-m}.
\end{aligned}$$

This proves, that $e_{cn}(D_n) \le c'/n$ for some two absolute constants $c, c' > 0$ and $n = 2^m, m \in \mathbb{N}$. The rest follows by monotonicity arguments.

[19] Let us first ignore the fact, that with this choice the numbers n_j are not natural.

Remark 11. The technical problem, that n_j must be natural numbers may be overcomed in two ways. The first is to take the integer part of $(1+\varepsilon)(m-j+1)2^{j-1}$. The second is to define $e_x(T)$ for all real $x > 1$ by $e_x(T) = e_{\lfloor x \rfloor}(T)$. We do not go into (rather technical) details.

Finally, we show, how we may use the entropy numbers of embeddings of function spaces to estimate the entropy numbers (and eigenvalues) of some differential operators.

Let $d : \operatorname{dom} d \subset L_2(-\pi,\pi) \to L_2(-\pi,\pi)$ be given by

$$\begin{aligned} &\operatorname{dom} d = \{f \in AC(-\pi,\pi) : f, f' \in L_2(-\pi,\pi), f(-\pi) = 0\} = \{f \in W_2^1(-\pi,\pi) : f(-\pi) = 0\},\\ &(df)(t) = f'(t), \quad t \in (-\pi,\pi). \end{aligned}$$

As d is unbouded, we estimate the entropy numbers of its (compact) inverse operator

$$(d^{-1}f)(x) = \int_{-\pi}^{x} f(t)dt, \quad d^{-1} : L_2(-\pi,\pi) \to \operatorname{dom}(d) \subset L_2(-\pi,\pi).$$

We use the following diagram

$$\begin{array}{ccccc} L_2(-\pi,\pi) & & \xrightarrow{d^{-1}} & & L_2(-\pi,\pi) \\ & \searrow_{d^{-1}} & & {}_{id}\nearrow & \\ & & W_2^1(-\pi,\pi) & & \end{array}$$

and get

$$e_n(d^{-1} : L_2 \to L_2) \le ||d^{-1} : L_2 \to W_2^1|| \cdot e_n(id : W_2^1 \to L_2) \le c/n, \quad n \in \mathbb{N}.$$

The estimate from below follows by

$$\begin{array}{ccccc} \{f \in W_2^1(-\pi,\pi) : f(-\pi) = 0\} & & \xrightarrow{id} & & L_2(-\pi,\pi) \\ & \searrow_{d} & & {}_{d^{-1}}\nearrow & \\ & & L_2(-\pi,\pi) & & \end{array}$$

and

$$c/n \le e_n(id : W_2^1 \to L_2) \le ||d : W_2^1 \to L_2|| \cdot e_n(d^{-1} : L_2 \to L_2), \quad n \in \mathbb{N}.$$

Combined with Carl-Triebel inequality, this implies the estimates of eigenvalues of d^{-1} from above, hence estimates of eigenvalues of d from below.

2 Approximation, Gelfand and Kolmogorov numbers

This section is devoted to other ways, how to measure and describe compactness. Unfortunately, there are too many of them and a detailed treatment would clearly go beyond the scope of this script. Therefore, we concentrate on three of them, which seem to be most useful in approximation theory, namely approximation, Gelfand and Kolmogorov numbers.

2.1 s-numbers

The general theory of s-numbers was created and developped by Pietsch in [4] and [5]. We quote (almost literarily) the Definition 2.2.1 from [5].

Definition 16. A rule
$$s : T \to \{s_n(T)\}_{n=1}^{\infty},$$
which assigns to every operator a scalar sequence, is said to be an *s-scale* if the following conditions are satisfied:

(i) $||T|\mathcal{L}(X,Y)|| = s_1(T) \geq s_2(T) \geq s_3(T) \geq \cdots \geq 0$ for $T \in \mathcal{L}(X,Y)$,

(ii) $s_{m+n-1}(S+T) \leq s_m(S) + s_n(T)$ for all $S, T \in \mathcal{L}(X,Y)$,

(iii) $s_n(T_1 T T_0) \leq ||T_0|| \cdot s_n(T) \cdot ||T_1||$ for $T_0 \in \mathcal{L}(X_0, X)$, $T \in \mathcal{L}(X,Y)$ and $T_1 \in \mathcal{L}(Y, Y_1)$,

(iv) if rank $T < n$, then $s_n(T) = 0$,

(v) $s_n(id : \ell_2^n \to \ell_2^n) = 1$.

Remark 12. (i) For $T \in \mathcal{L}(X,Y)$, we denote by rank T = dim $T(X)$ the dimension of its range. If rank $T < \infty$, then T is called finite dimensional.

(ii) Of course, in this definition are X and Y assumed to be Banach spaces. We generalize this definition to suit also quasi-Banach spaces. Also the property (iii) shall be slightly generalized.

(iii) It is the property (iv), what excludes the entropy numbers of being s-numbers. This lead Pietsch in [4, Chapter 12] to replace the axioms (i)-(v) with a different set of axioms leading to the so-called *pseudo s-functions*. We omit any details.

Definition 17. Let X, Y, Z be three quasi-Banach spaces and let Y be a p-Banach space with $0 < p \leq 1$. The rule $s : T \to \{s_n(T)\}_{n=1}^{\infty}$ is called an s-function, if

(i) $||T|\mathcal{L}(X,Y)|| \geq s_1(T) \geq s_2(T) \geq s_3(T) \geq \cdots \geq 0$ for all $T \in \mathcal{L}(X,Y)$,

(ii) $s_{m+n-1}^p(S+T) \leq s_m^p(S) + s_n^p(T)$ for all $S, T \in \mathcal{L}(X,Y)$,

(iii) $s_{m+n-1}(R \circ T) \leq s_m(R) \cdot s_n(T)$ for $R \in \mathcal{L}(Y,Z)$ and $T \in \mathcal{L}(X,Y)$,

(iv) if rank $T < n$, then $s_n(T) = 0$,

(v) $s_n(id : \ell_2^n \to \ell_2^n) = 1$.

Furthermore, we call $s_n(T)$ then n-th s-number of T.

2.2 Approximation numbers

The most important s-numbers are the *approximation numbers.*

Definition 18. Let X, Y be two quasi-Banach spaces and let $T \in \mathcal{L}(X,Y)$. Then we set
$$a_n(T) = \inf\{||T - A|\mathcal{L}(X,Y)|| : A \in \mathcal{L}(X,Y), \text{rank } A < n\}.$$

Theorem 19. *The approximation numbers a_n form an s-function.*

Proof. The proof of (i) is immediate. Let us just recall, that the operator $A \in \mathcal{L}(X,Y)$ defined as $Ax = 0$ for all $x \in X$ has the range $\{0\} \in Y$, which (according to the usual definition) has dimension 0.

Also the proof of (ii) follows by standard technique. Let $\varepsilon > 0$ and let $A, B \in \mathcal{L}(X,Y)$ be such, that

$$||S - A|\mathcal{L}(X,Y)|| \leq (1+\varepsilon)a_m(S), \quad ||T - B|\mathcal{L}(X,Y)|| \leq (1+\varepsilon)a_n(T).$$

Then

$$\begin{aligned} ||S - A + T - B|\mathcal{L}(X,Y)||^p &= \sup_{x\in B_X} ||(S-A)x + (T-B)x||_Y^p \\ &\leq \sup_{x\in B_X} ||(S-A)x||_Y^p + \sup_{x\in B_X} ||(T-B)x||_Y^p \\ &\leq ||S-A|\mathcal{L}(X,Y)||^p + ||T-B|\mathcal{L}(X,Y)||^p \\ &\leq (1+\varepsilon)^p \{a_m^p(S) + a_n^p(T)\} \end{aligned}$$

Finally, we recall from linear algebra, that rank $(A+B) \leq m-1+n-1$.

To prove (iii), let again $\varepsilon > 0$ and $A \in \mathcal{L}(X,Y), B \in \mathcal{L}(Y,Z)$ be such, that

$$||T - A|\mathcal{L}(X,Y)|| \leq (1+\varepsilon)a_n(T), \quad ||R - B|\mathcal{L}(Y,Z)|| \leq (1+\varepsilon)a_m(R).$$

Then

$$\begin{aligned} ||R\circ T - (R\circ A - B\circ A + B\circ T)|\mathcal{L}(X,Z)|| &= ||(R-B)\circ(T-A)|\mathcal{L}(X,Z)|| \\ \leq ||R-B|\mathcal{L}(Y,Z)||\cdot||T-A|\mathcal{L}(X,Y)|| &\leq (1+\varepsilon)^2 a_n(T)a_m(R). \end{aligned}$$

Let us also remark, that rank $[(R-B)\circ A + B\circ T] < m+n-1$.

The proof of (iv) is trivial, as well as the proof of

$$a_n(id : \ell_2^n \to \ell_2^n) \leq 1.$$

Finally, the lower estimate follows from the following Lemma. □

Lemma 20. *Let X be a quasi-Banach space with* $\dim(X) \geq n$. *Then* $a_n(id : X \to X) = 1$.

Proof. Let $a_n(id : X \to X) < 1$. Then there is an operator $A \in \mathcal{L}(X,X)$, such that

$$||id - A|\mathcal{L}(X,X)|| < 1 \quad \text{and} \quad \text{rank } A < n.$$

Then the Neumann series[20] of $A = id - (id - A)$ shows, that A must be invertible. Hence dim $X =$ rank $A < n$. □

Another interesting property is that a_n are actually the largest s-numbers.

Theorem 21. *The approximation numbers yield the largest s-function.*

Proof. Let $T \in \mathcal{L}(X,Y)$ and let $n \in \mathbb{N}$ be fixed. Then for every $\varepsilon > 0$ there is an operator $A \in \mathcal{L}(X,Y)$, such that

$$||T - A|\mathcal{L}(X,Y)|| < (1+\varepsilon)a_n(T) \quad \text{and} \quad \text{rank } A < n.$$

Then, for an arbitrary s-function s, it follows for suitable $0 < p \leq 1$

$$s_n^p(T) \leq ||T - A|\mathcal{L}(X,Y)||^p + s_n^p(A) \leq (1+\varepsilon)^p a_n^p(T).$$

□

[20] cf. Exercise 27

Remark 13. (i) It is very well known, that the identity on every infinite-dimensional Banach space is not compact. This is true also for quasi-Banach spaces - the reader may consult [1].

(ii) The relation between the set of all compact linear operators from X to Y (denoted by $\mathcal{K}(X,Y)$) and all continuous linear operators $\mathcal{L}(X,Y)$ may be very interesting. We quote (without proof) the *Pitt theorem*, which states that for $1 \le q < p < \infty$ the following identity is true: $\mathcal{L}(\ell_p, \ell_q) = \mathcal{K}(\ell_p, \ell_q)$. Hence, every bounded operator from ℓ_p into ℓ_q is also compact. Also $\mathcal{L}(c_0, \ell_q) = \mathcal{K}(c_0, \ell_q)$.

(iii) If $a_n(T) \to 0$, then T is compact, cf. Exercise 26. It was one of the most famous open problems, if also the converse is true. The counterexample was constructed by P. Enflo (and awarded live goose in Warsaw by Mazur). Nevertheless, the counterexample is very sophisticated and for "usual" spaces the converse is really true. Nevertheless, there are also exceptions. For example, if H is a separable Hilbert space, then $\mathcal{L}(H)$ does *not* have the approximation property.

Theorem 22. *Let X be a Banach space and Y a separable Banach space. If there is a sequence $\{S_n\}_{n=1}^\infty \subset \mathcal{L}(Y)$ of finite-dimensional operators, such that*

$$\lim_{n\to\infty} S_n y = y$$

for every $y \in Y$, then $\overline{\mathcal{F}(X,Y)} = \mathcal{K}(X,Y)$.

Proof. Let $T \in \mathcal{K}(X,Y)$. Then $S_nT \in \mathcal{F}(X,Y)$ and it is enough to show, that

$$||S_nT - T|\mathcal{L}(X,Y)|| \to 0.$$

According to the Theorem of Banach-Steinhaus, there is a $K \in \mathbb{R}$, such that $\sup_n ||S_n|\mathcal{L}(X,Y)|| \le K < \infty$.

Let $\varepsilon > 0$ be arbitrary and let $\{y_1, \dots, y_r\} \subset Y$ be such that

$$\overline{T(B_X)} \subset \bigcup_{j=1}^r y_j + \varepsilon B_Y.$$

Then there is an $N \in \mathbb{N}$, such that $||S_n y_j - y_j||_Y < \varepsilon$ for all $j = 1, \dots, r$ and all $n \ge N$. This leads to

$$||S_nTx - Tx||_Y \le ||S_n(Tx - y_j)||_Y + ||S_n y_j - y_j||_Y + ||y_j - Tx||_Y \le K\varepsilon + \varepsilon + \varepsilon = (K+2)\varepsilon$$

for all $x \in B_X$, all $n \ge N$ and appropriately chosen $j \in \{1, \dots, r\}$. □

Remark 14. (i) The assumption of this theorem is satisfied for example for the spaces $Y = c_0, Y = \ell_p, Y = L_p([0,1])$ with $1 \le p < \infty$ or $Y = C([0,1])$.

Theorem 23. *Let $0 < p \le \infty$ and let $\sigma = (\sigma_1, \sigma_2, \dots)$ with $\sigma_1 \ge \sigma_2 \ge \dots \ge 0$ be a non-increasing sequence. We define the diagonal operator D_σ as*

$$D_\sigma : \ell_p \to \ell_p, \quad D_\sigma x = (\sigma_1 x_1, \sigma_2 x_2, \dots).$$

Then

$$a_n(D_\sigma) = \sigma_n, \qquad n \in \mathbb{N}.$$

Proof. Step 1. Estimate from above

Let $n \in \mathbb{N}$. We denote by D_σ^{n-1} the $(n-1)$th sectional operator

$$D_\sigma^{n-1} x = (\sigma_1 x_1, \dots, \sigma_{n-1} x_{n-1}, 0, 0, \dots).$$

Then

$$a_n(D_\sigma) \le ||D_\sigma - D_\sigma^{n-1} | \mathcal{L}(\ell_p, \ell_p)|| = \sigma_n.$$

Step 2. Estimate from below

For the estimate from below, we use the following operators

$$\begin{aligned} D_\sigma^{(n)} : \ell_p^n \to \ell_p^n, &\quad D_\sigma^{(n)} x = (\sigma_1 x_1, \sigma_2 x_2, \dots, \sigma_{n-1} x_{n-1}, \sigma_n x_n), \\ I_p^n : \ell_p^n \to \ell_p^n, &\quad I_p^n(x_1, \dots, x_n) = (x_1, \dots, x_n), \\ J_n : \ell_p^n \to \ell_p, &\quad J_n(x_1, \dots, x_n) = (x_1, \dots, x_n, 0, 0, \dots), \\ P_n : \ell_p \to \ell_p^n, &\quad P_n(x_1, \dots, x_n, x_{n+1}, \dots) = (x_1, x_2, \dots, x_n). \end{aligned}$$

We may assume, that $\sigma_n \neq 0$ - otherwise, there is nothing to prove. Then we may calculate

$$\begin{aligned} 1 \le a_n(I_p^n) &= a_n((D_\sigma^{(n)})^{-1} \circ D_\sigma^{(n)}) \le ||(D_\sigma^{(n)})^{-1}|| \cdot a_n(D_\sigma^{(n)}) \\ &= \sigma_n^{-1} a_n(D_\sigma^{(n)}) = \sigma_n^{-1} a_n(P_n \circ D_\sigma \circ J_n) \le \sigma_n^{-1} \cdot ||P_n|| \cdot a_n(D_\sigma) \cdot ||J_n|| = \sigma_n^{-1} \cdot a_n(D_\sigma). \end{aligned}$$

□

2.3 Gelfand and Kolmogorov numbers

In this section we define the Gelfand and Kolmogorov numbers, prove their basic properties and study their relation to approximation numbers.

Definition 24. Let X, Y be two quasi-Banach spaces and let $T \in \mathcal{L}(X, Y)$.

(i) We define the n-th *Gelfand number* of the operator T as

$$c_n(T) = \inf\{||T \circ J_M^X | \mathcal{L}(M, Y)|| : M \subset X, \quad \text{codim } M < n\},$$

where $J_M^X : M \to X$ denotes the canonical embedding of a subspace $M \subset X$ into X.

(ii) We define the n-th *Kolmogorov number* of the operator T as

$$d_n(T) = \inf\{||Q_N^Y \circ T | \mathcal{L}(X, Y/N)|| : N \subset Y, \quad \dim N < n\},$$

where $Q_N^Y : Y \to Y/N$ is the quotient map.

Remark 15. (i) On many occasions, one uses an equivalent definition of c_n, namely

$$c_n(T) = \inf_{\substack{M \subset X \\ \text{codim } M < n}} \sup_{\substack{x \in M \\ ||x||_X \le 1}} ||Tx||_Y.$$

(ii) The quotient space Y/N is the space of cosets $\overline{y} = \{y - z : z \in N\}$ equipped with the quotient (quasi-)norm

$$||\overline{y}|Y/N|| = \inf\{||y - z||_Y : z \in N\}.$$

The quotient map Q_N^Y is defined as $Q_N^Y y = \overline{y}$.

Using this terminology, we may rewrite the definition of Kolmogorov numbers as

$$d_n(T) \;=\; \inf_{\substack{N\subset Y\\ \dim\ N<n}} \sup_{\substack{x\in X\\ ||x||_X\le 1}} ||\overline{Tx}|Y/N|| = \inf_{\substack{N\subset Y\\ \dim\ N<n}} \sup_{\substack{x\in X\\ ||x||_X\le 1}} \inf_{z\in N} ||Tx-z||_Y.$$

(iii) Observe, that all the structures used so far (canonical embedding, quotient space, quotient mapping) are well definied also for quasi-Banach spaces. Nevertheless, if X and/or Y are Banach spaces, then there are numerous equivalent definitions of Gelfand and Kolmogorov numbers. We shall see some of them later on.

(iv) The definition of approximation numbers was based on a linear approximation of an operator T on the whole space X. Both the Gelfand as well as the Kolmogorov numbers involve certain nonlinearity (which shall be discussed later in the Exercises) and therefore the following proposition should not be really surprising.

Proposition 25. *Let X and Y be two quasi-Banach spaces and let $T\in\mathcal{L}(X,Y)$. Then*

$$c_n(T)\le a_n(T) \quad \textit{and} \quad d_n(T)\le a_n(T) \quad \textit{for all} \quad n\in\mathbb{N}.$$

Proof. Let $\varepsilon>0$ be arbitrary and let $A\in\mathcal{L}(X,Y)$ be such, that $||T-A|\mathcal{L}(X,Y)||<(1+\varepsilon)a_n(T)$ and rank $A<n$.

Step 1. $d_n(T)\le a_n(T)$.

We put $N=A(X)$. Then dim $N=$ rank $A<n$ and

$$d_n(T)\le \sup_{x\in B_X}\inf_{y\in N}||Tx-y||_Y\le \sup_{x\in B_X}||Tx-Ax||_Y=||T-A|\mathcal{L}(X,Y)||\le(1+\varepsilon)a_n(T).$$

Step 2. $c_n(T)\le a_n(T)$.

We put $M=\text{kern } A=\{x\in X: Ax=0\}$. Then we have

$$c_n(T)\le \sup_{\substack{x\in\text{kern } A\\ ||x||_X\le 1}}||Tx||_Y=\sup_{\substack{x\in\text{kern } A\\ ||x||_X\le 1}}||Tx-Ax||_Y\le||T-A|\mathcal{L}(X,Y)||\le(1+\varepsilon)a_n(T).$$

The last thing, which one has to consider is to show, that codim $M<n$. We postpone this (rather algebraic) proof to the Exercise 28. □

Theorem 26. *The Gelfand numbers as well as the Kolmogorov numbers form an s-function.*

Proof. Step 1. **Gelfand numbers**

The proof of (i) follows from the observation, that the only space $M\subset X$ with codim $M<1$ is the space X itself. And then

$$c_1(T)=\sup_{x\in B_X}||Tx||_Y=||T|\mathcal{L}(X,Y)||.$$

The proof of the property $c_j(T)\ge c_{j+1}(T)$ for all $j\in\mathbb{N}$ is trivial.

To prove (ii) we take $\varepsilon>0$ arbitrary and find

$$\begin{aligned} M_1\subset X \quad &\text{with}\quad \text{codim } M_1<m: x\in M_1 \implies ||Sx||_Y\le(1+\varepsilon)c_m(S)||x||_X,\\ M_2\subset X \quad &\text{with}\quad \text{codim } M_2<n: x\in M_2 \implies ||Tx||_Y\le(1+\varepsilon)c_n(T)||x||_X.\end{aligned}$$

Then we obtain

$$||(S+T)x||_Y^p \le ||Sx||_Y^p + ||Tx||_Y^p \le (1+\varepsilon)^p ||x||_X^p \left(c_m^p(S) + c_n^p(T)\right)$$

for all $x \in M_1 \cap M_2$ - a subspace of X with codimension smaller then $m+n-1$.

The proof of (iii) follows similarly.

$$\begin{aligned} M_1 \subset X \quad &\text{with} \quad \text{codim } M_1 < n : x \in M_1 \implies ||Tx||_Y \le (1+\varepsilon) c_n(T) ||x||_X, \\ M_2 \subset Y \quad &\text{with} \quad \text{codim } M_2 < m : y \in M_2 \implies ||Ry||_Z \le (1+\varepsilon) c_m(R) ||y||_Y. \end{aligned}$$

Then

$$||R(Tx)||_Z \le (1+\varepsilon) c_m(R) ||Tx||_Y \le (1+\varepsilon)^2 c_m(R) c_n(T) ||x||_X$$

for all $x \in X$ with $x \in M_1$ and $Tx \in M_2$, i.e. for all $x \in M_1 \cap T^{-1}(M_2)$ - a subspace of X with codimension smaller then $m+n-1$, cf. Exercise 28.

To (iv): if rank $T < n$, then codim kern $T < n$ and $T|_{\text{kern } T} = 0$, hence $c_n(T) = 0$.

Finally (v) follows from Lemma 27.

Step 2. **Kolmogorov numbers**

The only subspace $N \subset Y$ with dim $N < 1$ is the space $\{0\} \subset Y$. Hence

$$d_1(T) = \sup_{x \in B_X} ||Tx - 0||_Y = ||T|\mathcal{L}(X,Y)||.$$

The inequality $d_j(T) \ge d_{j+1}(T), j \in \mathbb{N}$ is again trivial and the proof of (i) is therefore complete.

To prove (ii) we take $\varepsilon > 0$ arbitrary and find

$$\begin{aligned} N_1 \subset Y \quad &\text{with} \quad \dim N_1 < m : x \in X \implies \exists y_1 \in N_1 : ||Sx - y_1||_Y \le (1+\varepsilon) d_m(S) ||x||_X, \\ N_2 \subset Y \quad &\text{with} \quad \dim N_2 < n : x \in X \implies \exists y_2 \in N_2 : ||Tx - y_2||_Y \le (1+\varepsilon) d_n(T) ||x||_X. \end{aligned}$$

Take $x \in X$ and find corresponding y_1 and y_2 as described above. Then we obtain

$$||(S+T)x - y_1 - y_2||_Y^p \le ||Sx - y_1||_Y^p + ||Tx - y_2||_Y^p \le (1+\varepsilon)^p ||x||_X^p \left(d_m^p(S) + d_n^p(T)\right).$$

Here, $y = y_1 + y_2 \in N_1 + N_2$ - a subspace of Y with dimension smaller then $m+n-1$.

The proof of (iii) follows similarly.

$$\begin{aligned} N_1 \subset Y \quad &\text{with} \quad \dim N_1 < n : x \in X \implies \exists \overline{y} \in N_1 : ||Tx - \overline{y}||_Y \le (1+\varepsilon) d_n(T) ||x||_X, \\ N_2 \subset Z \quad &\text{with} \quad \dim N_2 < m : y \in Y \implies \exists z \in N_2 : ||Ry - z||_Z \le (1+\varepsilon) d_m(R) ||y||_Y. \end{aligned}$$

Let us take $x \in X$. We find $\overline{y}$ to Tx as described above and z to $R(Tx - \overline{y})$ instead of Ry. Then we may estimate

$$||R(Tx) - R(\overline{y}) - z||_Z = ||R(Tx - \overline{y}) - z||_Z \le (1+\varepsilon) d_m(R) ||Tx - \overline{y}||_Y \le (1+\varepsilon)^2 d_m(R) d_n(T) ||x||_X,$$

where $R(\overline{y}) + z \in R(N_1) + N_2$ - a subspace of Z with dimension smaller then $m+n-1$.

To (iv): if rank $T < n$, then dim $T(X) < n$

$$d_n(T) \le \sup_{x \in B_X} \inf_{y \in T(x)} ||Tx - y||_Y = 0.$$

Finally (v) follows from Lemma 27. □

Lemma 27. *Let X be a quasi-Banach space with* $\dim(X) \geq n$. *Then* $c_n(id : X \to X) = d_n(id : X \to X) = 1$.

Proof. Let $n \in \mathbb{N}$ and let X be a space with $\dim X \geq n$.

Step 1. Gelfand numbers

Let $M \subset X$ be a subspace of X with codimension smaller then n. Then $M \neq \{0\}$. Hence

$$c_n(id : X \to X) = \inf_{\substack{M \subset X \\ \text{codim } M < n}} \sup_{\substack{x \in M \\ ||x||_X \leq 1}} ||id(x)||_Y \geq 1.$$

Step 2. Kolmogorov numbers

Let $\varepsilon > 0$ and let $N \subset X$ with $\dim N < n$. Then $N \neq X$ and accordning to the Riesz's lemma, there is an $x_{N,\varepsilon} \in X \setminus N$, such that $||x_{N,\varepsilon}||_X = 1$ and $||x_{N,\varepsilon} - y||_X \geq \frac{1}{1+\varepsilon}$ for all $y \in N$. Hence

$$d_n(T) \geq \inf_{\substack{N \subset Y \\ \dim N < n}} \inf_{y \in N} ||Tx_{N,\varepsilon} - y|| \geq \frac{1}{1+\varepsilon}.$$

□

Theorem 28. *a) Let X and Y be two Banach spaces and let $T \in \mathcal{L}(X, Y)$. Then T is compact if, and only if, $c_n(T) \to 0$,*

b) Let X and Y be two quasi-Banach spaces and let $T \in \mathcal{L}(X, Y)$. Then T is compact if, and only if, $d_n(T) \to 0$.

Proof. Step 1. Gelfand numbers

Let T be compact, i.e. for every $\varepsilon > 0$

$$T(B_X) \subset \bigcup_{j=1}^{J} y_j + \varepsilon B_Y \tag{2.1}$$

for suitable $J \in \mathbb{N}$ and $\{y_1, \dots, y_J\} \in Y$.

According to the Hahn-Banach Theorem, there are functionals $\beta_j \in Y'$, such that

$$|\beta_j(y_j)| = ||y_j||_Y \quad \text{and} \quad ||\beta_j||_{Y'} = 1, \quad j = 1, \dots, J.$$

We define $\alpha_j \in X'$ by $\alpha_j(x) = \beta_j(T(x)), j = 1, \dots, J$ and $M = \{x \in X : \alpha_j(x) = 0 \text{ for all } j = 1, \dots, J\}$. Then,[21]

for $x \in M$ with $||x||_X \leq 1$ and an appropriate $j \in \{1, \dots, J\}$,

$$||Tx||_Y \leq ||Tx - y_j||_Y + ||y_j||_Y \leq \varepsilon + |\beta_j(y_j)| \leq \varepsilon + |\beta_j(Tx - y_j)| + |\beta_j(Tx)| \leq \varepsilon + ||\beta_j||_{Y'} \cdot ||Tx - y_j||_Y \leq 2\varepsilon.$$

Conversely, let $c_n(T) \to 0$. Then for every $\varepsilon > 0$, there are $J \in \mathbb{N}$ and $\alpha_1, \dots, \alpha_J \in X'$ such that for every

$$x \in M = \bigcap_{j=1}^{J} \text{kern } \alpha_j = \{x \in X : \alpha_1(x) = \dots = \alpha_J(x) = 0\}$$

[21] A detailed investigation of the next line show, that $||y_j||_Y < \varepsilon$, which seems to be in contradiction with (2.1), but this holds only for those y_j's, which play a role in covering of $T(M)$.

the following inequality holds

$$||Tx||_Y \le \varepsilon ||x||_X.$$

We show, that this implies, that T' is compact (and hence also T).

Let $\beta \in B_{Y'}$. Then $T'\beta \in X'$ and (according to the Hahn-Banach Theorem), there is a functional $\theta \in X'$ such that $\theta(x) = T'\beta(x)$ for every $x \in M$ and

$$||\theta||_{X'} = \sup_{x \in B_X} |\theta(x)| = \sup_{\substack{x \in M \\ ||x||_X \le 1}} |\theta(x)| = \sup_{\substack{x \in M \\ ||x||_X \le 1}} |(T'\beta)(x)| = \sup_{\substack{x \in M \\ ||x||_X \le 1}} |\beta(Tx)| \le ||\beta||_{Y'} \cdot \sup_{\substack{x \in M \\ ||x||_X \le 1}} ||Tx|| \le \varepsilon.$$

We define $\tilde{M} = Lin(\alpha_1, \dots, \alpha_J) \subset X'$. As $(\theta - T'\beta)(x) = 0$ for all $x \in M$, we conclude, that $\theta - T'\beta \in \tilde{M}$. Hence, $T'\beta = (T'\beta - \theta) + \theta \in \tilde{M} + \varepsilon B_{X'}$.

This holds for all $\beta \in B_{Y'}$, hence

$$T'(B_{Y'}) \subset \tilde{M} + \varepsilon B_{X'}.$$

But if $T'\beta = \tilde{m} + \varepsilon\chi$ with $\tilde{m} \in \tilde{M}$ and $\chi \in B_{X'}$, then

$$||\tilde{m}||_{X'} \le ||T'\beta||_{X'} + \varepsilon ||\chi||_{X'} \le ||T'|\mathcal{L}(Y', X')|| \cdot ||\beta||_{Y'} + \varepsilon.$$

Hence even

$$T'(B_{Y'}) \subset ||T'|\mathcal{L}(Y', X')|| B_{X'} \cap \tilde{M} + \varepsilon B_{X'}.$$

The set $||T'|\mathcal{L}(Y', X')|| B_{X'} \cap \tilde{M}$ is a bounded set in a finite-dimensional space $\tilde{M}$ and may be covered like

$$||T'|\mathcal{L}(Y', X')|| B_{X'} \cap \tilde{M} \subset \bigcup_{k=1}^{K} \gamma_k + \varepsilon B_{X'},$$

which leads to

$$T'(B_{Y'}) \subset \bigcup_{k=1}^{K} \gamma_k + 2\varepsilon B_{X'}.$$

Step 2. Kolmogorov numbers

Let T be compact, i.e. for every $\varepsilon > 0$

$$T(B_X) \subset \bigcup_{j=1}^{J} y_j + \varepsilon B_Y$$

for suitable $J \in \mathbb{N}$ and $\{y_1, \dots, y_J\} \in Y$. Put $N = Lin(y_1, \dots, y_J)$.

Then

$$d_{J+1}(T) \le \sup_{x \in B_X} \inf_{y \in N} ||Tx - y||_Y \le \sup_{x \in B_X} \inf_{j=1,\dots,J} ||Tx - y_j||_Y \le \varepsilon.$$

Let on the other hand $d_n(T) \to 0$. Then for every $\varepsilon > 0$ there is a finite-dimensional subspace $N \subset Y$, such that

$$\sup_{x \in B_X} \inf_{y \in N} ||Tx - y||_Y < \varepsilon.$$

This may be rewritten as

$$T(B_X) \subset N + \varepsilon B_Y.$$

But if $Tx = y + \varepsilon z$ with $y \in N$ and $z \in B_Y$, then

$$||y||_Y^p \leq ||Tx||_Y^p + \varepsilon^p$$

for a suitable $0 < p \leq 1$. Hence also

$$T(B_X) \subset N \cap (||T|\mathcal{L}(X,Y)||^p + \varepsilon^p)^{1/p} B_Y + \varepsilon B_Y.$$

The set $N \cap (||T|\mathcal{L}(X,Y)||^p + \varepsilon^p)^{1/p} B_Y$ is a bounded set in finite-dimensional quasi-Banach space N and may be therefore covered

$$N \cap (||T|\mathcal{L}(X,Y)||^p + \varepsilon^p)^{1/p} B_Y \subset \bigcup_{j=1}^m z_j + \varepsilon B_Y$$

for appropriate $m \in \mathbb{N}$ and $z_1, \dots, z_m \in N$. Hence

$$T(B_X) \subset \bigcup_{j=1}^m z_j + 2^{1/p-1} \varepsilon B_Y$$

and T is compact. □

When dealing with Hilbert spaces, the situation is usually much simpler - some (or even all) of the s-numbers coincide.

Theorem 29. *Let X and Y be two quasi-Banach spaces and let $T \in \mathcal{L}(X,Y)$.*

a) If X is even a Hilbert space, then $c_n(T) = a_n(T)$.

b) If Y is even a Hilbert space, then $d_n(T) = a_n(T)$.

c) If both X and Y are Hilbert spaces, then $c_n(T) = d_n(T) = a_n(T)$.

Proof. Of course, c) is a simple corollary of a) and b).

Step 1. Gelfand numbers

Let $n \in \mathbb{N}$ and $\varepsilon > 0$ be arbitrary. Then there is a subspace $M \subset X$ with codim $M < n$, such that

$$x \in M \implies ||Tx||_Y \leq (1+\varepsilon) c_n(T) ||x||_X.$$

We define the space $M^\perp = \{y \in X :< x, y >_X = 0 \text{ for all } x \in M\}$ of all elements orthogonal to the all elements of M. Finally, we denote by $P_{M^\perp}$ the orthogonal projection of X onto $M^\perp$ and set $A = T \circ P_{M^\perp}$.

Then

$$\begin{aligned} a_n(T) &\leq ||T - A|\mathcal{L}(X,Y)|| = \sup_{x \in B_X} ||Tx - Ax||_Y = \sup_{x \in B_X} ||Tx - T(P_{M^\perp} x)||_Y \\ &= \sup_{x \in B_X} ||T(\underbrace{x - P_{M^\perp} x}_{\in M})||_Y \leq (1+\varepsilon) c_n(T) \sup_{x \in B_X} ||x - P_{M^\perp} x||_X \leq (1+\varepsilon) c_n(T). \end{aligned}$$

Finally, we let $\varepsilon \to 0$.

Step 2. Kolmogorov numbers

Let $n \in \mathbb{N}$ and $\varepsilon > 0$ be arbitrary. Then there is a subspace $N \subset Y$ with $\dim N < n$, such that

$$x \in X \implies \inf_{y \in N} ||Tx - y||_Y \le (1+\varepsilon) d_n(T) ||x||_X.$$

We set $A = P_N T$, where P_N is again the orthogonal projection, this time of Y onto N. Then

$$\begin{aligned} a_n(T) &\le ||T - A|\mathcal{L}(X,Y)|| = \sup_{x \in B_X} ||Tx - Ax||_Y = \sup_{x \in B_X} ||Tx - P_N(Tx)||_Y \\ &= \sup_{x \in B_X} \inf_{y \in N} ||Tx - y||_Y \le (1+\varepsilon) d_n(T) \end{aligned}$$

and we let again $\varepsilon \to 0$. □

Also the relation of interpolation theory to Gelfand and Kolmogorov numbers is easily seen.

Theorem 30. *Let X, T, $Y_0, Y_\theta, Y_1 \subset \mathbb{Y}$, $0 < p \le 1$ and $0 < \theta < 1$ satisfy all the assumptions of Theorem 10. Then*

$$c_{n_0+n_1-1}(T : X \to Y_\theta) \le c_{n_0}^{1-\theta}(T : X \to Y_0) \cdot c_{n_1}^{\theta}(T : X \to Y_1), \quad n_0, n_1 \in \mathbb{N}.$$

Proof. Let $M_0, M_1 \subset X$ be two subspaces of X with

$$\begin{aligned} &\operatorname{codim} M_0 < n_0 \quad \text{and} \quad x \in M_0 \implies ||Tx||_{Y_0} \le (1+\varepsilon) c_{n_0}(T : X \to Y_0) ||x||_X, \\ &\operatorname{codim} M_1 < n_1 \quad \text{and} \quad x \in M_1 \implies ||Tx||_{Y_1} \le (1+\varepsilon) c_{n_1}(T : X \to Y_1) ||x||_X. \end{aligned}$$

Put $M = M_0 \cap M_1$. Then $\operatorname{codim} M < n_0 + n_1 - 1$ and for $x \in M$

$$||Tx||_{Y_\theta} \le ||Tx||_{Y_0}^{1-\theta} \cdot ||Tx||_{Y_1}^{\theta} \le (1+\varepsilon) ||x||_X c_{n_0}^{1-\theta}(T : X \to Y_0) c_{n_1}^{\theta}(T : X \to Y_1)$$

and the result follows. □

Theorem 31. *Let $X_0, X_\theta, X_1 \subset \mathbb{X}$, T, Y, $0 < p \le 1$ and $0 < \theta < 1$ satisfy all the assumptions of Theorem 11. Then*

$$d_{n_0+n_1-1}(T : X_\theta \to Y) \le 2^{1/p} d_{n_0}^{1-\theta}(T : X_0 \to Y) \cdot d_{n_1}^{\theta}(T : X_1 \to Y), \quad n_0, n_1 \in \mathbb{N}.$$

Proof. Let us abbreviate $d_{n_0} = d_{n_0}(T : X_0 \to Y)$ and $d_{n_1} = d_{n_1}(T : X_1 \to Y)$

Let $N_0, N_1 \subset Y$ be two subspaces of Y with

$$\begin{aligned} &\dim N_0 < n_0 \quad \text{and} \quad x \in X_0 \implies \inf_{y \in N_0} ||Tx - y||_Y \le (1+\varepsilon) d_{n_0} ||x||_{X_0}, \\ &\dim N_1 < n_1 \quad \text{and} \quad x \in X_1 \implies \inf_{y \in N_1} ||Tx - y||_Y \le (1+\varepsilon) d_{n_1} ||x||_{X_1}. \end{aligned}$$

Put $N = N_0 + N_1$. Then $\dim N < n_0 + n_1 - 1$. Let $x \in X$ and $t > 0$ to be chosen later on. Then we find $x_0 \in X_0$ and $x_1 \in X_1$ such that

$$x = x_0 + x_1 \quad \text{and} \quad ||x_0||_{X_0} + t||x_1||_{X_1} \le t^\theta ||x||_{X_\theta}$$

and $y_0 \in N_0$ and $y_1 \in N_1$, such that

$$||Tx_0 - y_0||_Y \le (1+\varepsilon) d_{n_0} ||x_0||_{X_0} \quad \text{and} \quad ||Tx_1 - y_1||_Y \le (1+\varepsilon) d_{n_1} ||x_1||_{X_1}.$$

Finally, we arrive at

$$\begin{aligned} ||Tx - (y_0 + y_1)||_Y^p &\le ||Tx_0 - y_0||_Y^p + ||Tx_1 - y_1||_Y^p \le (1+\varepsilon)(d_{n_0}^p ||x_0||_{X_0}^p + d_{n_1}^p ||x_1||_{X_1}^p) \\ &\le (1+\varepsilon)(d_{n_0}^p t^{p\theta} ||x||_{X_\theta}^p + d_{n_1}^p t^{p(\theta-1)} ||x||_{X_\theta}^p) \le 2 d_{n_0}^{p(1-\theta)} d_{n_1}^{p\theta} ||x||_{X_\theta}^p, \end{aligned}$$

where we have chosen $t = d_{n_1}/d_{n_0}$. From this, the result follows. □

2.3.1 Duality

In this section, we consider the relation between $c_n(T)$ and $d_n(T')$ (and briefly also between $a_n(T)$ and $a_n(T')$). Some of the main ideas were already hidden in the proof of Theorem 28, but here we are going to discuss the concept of duality and its connection to s-numbers in detail.

In this whole section we assume X and Y to be Banach spaces, so that we can

- work with X' and Y',
- apply the Hahn-Banach theorem,
- consider the dual operator $T' \in \mathcal{L}(Y', X')$ for every $T \in \mathcal{L}(X, Y)$.

Let us recall the basic facts from functional analysis about duality.

If X and Y are two Banach spaces and $T \in \mathcal{L}(X, Y)$, then we define $T' \in \mathcal{L}(Y', X')$ by

$$[T'(\varphi)](x) = \varphi(T(x))$$

for every $\varphi \in Y'$ and $x \in X$. It is quite simple to see, that with this definition the operator T' is really linear and bounded (even with $||T|\mathcal{L}(X,Y)|| = ||T'|\mathcal{L}(Y',X')||$). According to *Schauder's theorem* we know, that T is compact if, and only if, T' is compact.

If X is a Banach space and X' is its dual space, then we denote by X'' the *second dual space*. The *canonical embedding* of X into X'' is defined as

$$\varepsilon_X : X \to X'', \qquad \varepsilon_X(x)(\varphi) = \varphi(x)$$

for every $x \in X$ and $\varphi \in X'$. If ε_X is even an isomporhism of X onto X'', then X is called *reflexive*.

Definition 32. Let X be a Banach space and X' be its dual space.

a) If $L \subset X$ is a subspace of X, we define

$$L^\perp = \{\varphi \in X' : \varphi(x) = 0 \text{ for all } x \in L\}.$$

b) If $M \subset X'$ is a subspace of X', we define

$$M_\perp = \{x \in X : \varphi(x) = 0 \text{ for all } \varphi \in M\}.$$

Both $L^\perp$ and $M_\perp$ are called *annihilators*.

Theorem 33. *Let X, X', L and M be as above. Then*

$$(L^\perp)_\perp = \overline{L}, \qquad (M_\perp)^\perp = \overline{M}^{\,\omega^*}.$$

Here $\overline{L}$ stands for the closure of L and $\overline{M}^{\,\omega^*}$ is the so-called weak-star closure of M.

Proof. Step 1. $(L^\perp)_\perp = \overline{L}$.

a) Let $x \in \overline{L}$ and let $\varphi \in L^\perp$. As $\varphi(L) = 0$ and φ is continuous, then also $\varphi(x) = 0$. Hence $\varphi(x) = 0$ for all $\varphi \in L^\perp$, which means, that $x \in (L^\perp)_\perp$.

b) Let $x \notin \overline{L}$. According to the Hahn-Banach Theorem, there is a $\varphi \in X'$, such that $\varphi(\overline{L}) = 0$ and $\varphi(x) \neq 0$. Hence, $x \notin (L^\perp)_\perp$.

Step 2. $L^\perp$ is weak-star closed in X' for all subspaces $L \subset X$.

Let us recall, that the set $\mathcal{O} \subset X'$ is weak-star open if, and only if,

$$\forall \varphi \in \mathcal{O}\ \exists \varepsilon > 0\ \exists x_1, \dots, x_n \in X : \{\psi \in X' : |\psi(x_i) - \varphi(x_i)| < \varepsilon \text{ for all } i = 1, \dots, n\} \subset \mathcal{O}.$$

Of course, a set $\mathcal{C} \subset X'$ is weak-star closed if, and only if, its complement is weak-star open, i.e.

$$\forall \varphi \notin \mathcal{C}\ \exists \varepsilon > 0\ \exists x_1, \dots, x_n \in X : \{\psi \in X' : |\psi(x_i) - \varphi(x_i)| < \varepsilon \text{ for all } i = 1, \dots, n\} \cap \mathcal{C} = \emptyset.$$

So, let us take $\varphi \notin L^\perp$. Then there is $x_0 \in L$, such that $0 < |\varphi(x_0)| =: \varepsilon$ and we observe that

$$\{\psi \in X' : |\psi(x_0) - \varphi(x_0)| < \varepsilon\} \cap L^\perp = \emptyset.$$

Hence, $L^\perp$ is weak-star closed.

Step 3. $\overline{M}^{\,\omega^*} \subset (M_\perp)^\perp$

Obviuously, $M \subset (M_\perp)^\perp$. But the later set is weak-star closed (cf. Step 2.) and the inclusion follows.

Step 4. $(M_\perp)^\perp \subset \overline{M}^{\,\omega^*}$

Let $\varphi \notin \overline{M}^{\,\omega^*}$. Then we may apply the Hahn-Banach Theorem with respect to the weak-star topology on X' and find $\Psi \in (X', \omega^*)'$, such that $\Psi(\varphi) \neq 0$ and $\Psi(\overline{M}^{\,\omega^*}) = 0$. It is a standard fact from functional analysis, that $(X', \omega^*)' = \varepsilon_X(X) \subset X''$. Hence Ψ may be represented by an $x \in X$ with $\varphi(x) \neq 0$ and $\psi(x) = 0$ for all $\psi \in \overline{M}^{\,\omega^*}$ - especially $x \in M_\perp$. Hence $\varphi \notin (M_\perp)^\perp$. □

The relation between annihilators and approximation theory is given in the following fundamental Lemma.

Lemma 34. *a) Let X be a Banach space, $x \in X$ and let $L \subset X$ be a subspace of X. Then*

$$\inf_{y \in L} ||x - y||_X = \max\{|\varphi(x)| : \varphi \in L^\perp, ||\varphi||_{X'} \le 1\}. \tag{2.2}$$

b) Let X be a Banach space, X' its dual space, let $M \subset X'$ be a weak-star closed subspace and let $\varphi \in X'$. Then

$$\min_{\psi \in M} ||\varphi - \psi||_{X'} = \sup\{|\varphi(x)| : x \in M_\perp, ||x||_X \le 1\}. \tag{2.3}$$

Proof. a) We may assume, that L is closed - i.e. both sides of (2.2) do not change, if we replace L by $\overline{L}$. If $x \in L$, then there is nothing to prove. So, we may also assume, that $x \notin L$.

Let $\tilde{L} = lin\{L, x\} \subset X$. It means, that every element $\omega \in \tilde{L}$ may be written (in a unique way) as $\omega = \lambda x + y$, where $\lambda \in \mathbb{R}$ and $y \in L$.

Let $\varphi : \tilde{L} \to \mathbb{R}$ be linear and continuous, such that $\varphi(L) = 0$, $\varphi(x) > 0$ and $||\varphi||_{(\tilde{L})'} = 1$.

Then for every $\varepsilon > 0$ there is $z_\varepsilon \in \tilde{L}$, such that

$$||z_\varepsilon||_X = 1 \quad \text{and} \quad \varphi(z_\varepsilon) \ge (1 - \varepsilon)||\varphi||_{(\tilde{L})'} = 1 - \varepsilon.$$

We decompose $z_\varepsilon = \lambda_\varepsilon x + y_\varepsilon$. Then $\varphi(z_\varepsilon) = \lambda_\varepsilon \varphi(x) + \varphi(y_\varepsilon) = \lambda_\varepsilon \varphi(x) > 1 - \varepsilon$.

$$\left\| x + \frac{y_\varepsilon}{\lambda_\varepsilon} \right\|_X = \left\| \frac{z_\varepsilon}{\lambda_\varepsilon} \right\|_X = \frac{1}{|\lambda_\varepsilon|} \cdot ||z_\varepsilon||_X = \frac{1}{\lambda_\varepsilon} \le \frac{\varphi(x)}{1 - \varepsilon}.$$

And the result follows by

$$\inf_{y\in L} ||x-y||_X \le \inf_{0<\varepsilon<1} \left\| x + \frac{y_\varepsilon}{\lambda_\varepsilon} \right\| \le \inf_{0<\varepsilon<1} \frac{\varphi(x)}{1-\varepsilon} = \varphi(x).$$

To prove the second inequality in (2.2), we calculate for arbitrary $y \in L$:

$$||x-y||_X = \sup_{\varphi\in B_{Y'}} |\varphi(x-y)| \ge \sup_{\substack{\varphi\in L^\perp\\ ||\varphi||_{Y'}\le 1}} |\varphi(x-y)| = \sup_{\substack{\varphi\in L^\perp\\ ||\varphi||_{Y'}\le 1}} |\varphi(x)|.$$

Let us remark, that it also follows from the proof, that the maximum on the right hand side of (2.2) is attained - namely by the φ constructed above.

b) Let $M \subset X'$, $\varphi \in X'$ and let $M \subset X'$ be a weak-star closed subspace of X'.

It follows from Theorem 33, that

$$\theta \in X' : \theta \in M \Leftrightarrow \theta(M_\perp) = 0.$$

We put

$$a := \sup\{|\varphi(x)| : x \in M_\perp, ||x||_X \le 1\} < \infty.$$

Then

$$||\varphi-\psi||_{X'} = \sup_{x\in B_X} |\varphi(x)-\psi(x)| \ge \sup_{\substack{x\in M_\perp\\ ||x||_X\le 1}} |\varphi(x)| = a$$

holds for all $\psi \in M$ and hence also for the infimum.

On the other hand, we apply Hahn-Banach Theorem to obtain $\psi \in X'$ with

$$\psi(x) = \varphi(x) \text{ for all } x \in M \quad \text{and} \quad ||\psi||_{X'} = a.$$

We put $\theta = \varphi - \psi$. Then $\theta(M_\perp) = 0$ and hence $\theta \in M$. But also

$$||\varphi-\theta||_{X'} = ||\psi||_{X'} = a.$$

Hence the minimum on the left-hand side of (2.3) is attained (in θ). □

Theorem 35. *Let X, Y be two Banach spaces and $T \in \mathcal{L}(X,Y)$.*

a) Then

$$c_n(T) = d_n(T') \tag{2.4}$$

and

$$c_n(T') \le d_n(T). \tag{2.5}$$

b) If T is even compact, then

$$c_n(T') \ge d_n(T). \tag{2.6}$$

Proof. Step 1. Proof of (2.4)

Let $M \subset X$ be a subspace of X with codim $X < n$. We set $N = M^\perp$. Then $N_\perp = (M^\perp)_\perp = M$ and

$$\sup_{\substack{x\in M\\ ||x||_X\le 1}} ||Tx||_Y = \sup_{\substack{x\in M\\ ||x||_X\le 1}} \sup_{\substack{\varphi\in Y'\\ ||\varphi||_{Y'}\le 1}} |\varphi(Tx)| = \sup_{\substack{\varphi\in Y'\\ ||\varphi||_{Y'}\le 1}} \sup_{\substack{x\in M\\ ||x||_X\le 1}} |(T'\varphi)(x)|$$

$$= \sup_{\substack{\varphi\in Y'\\ ||\varphi||_{Y'}\le 1}} \sup_{\substack{x\in N_\perp\\ ||x||_X\le 1}} |(T'\varphi)(x)| = \sup_{\substack{\varphi\in Y'\\ ||\varphi||_{Y'}\le 1}} \sup_{\psi\in N} ||T'\varphi-\psi||_{X'}.$$

Taking infimum over all M finishes the proof.

Step 2. Proof of (2.5)

Let us recall, that

$$d_n(T) = \inf_{\substack{N\subset Y\\ \dim\ N<n}} \sup_{x\in B_X} \inf_{y\in N} ||Tx-y||_Y,$$
$$c_n(T') = \inf_{\substack{M\subset Y'\\ \mathrm{codim}\ M<n}} \sup_{\substack{\varphi\in M\\ ||\varphi||_{Y'}\le 1}} ||T'\varphi||_{X'} = \inf_{\substack{M\subset Y'\\ \mathrm{codim}\ M<n}} \sup_{\substack{\varphi\in M\\ ||\varphi||_{Y'}\le 1}} \sup_{x\in B_X} |\varphi(Tx)|.$$

Now let $\varepsilon>0$. Then there is a subspace $N\subset Y$ with $\dim N<n$, such that

$$\forall x\in B_X\ \exists\, y\in N : ||Tx-y||_Y \le (1+\varepsilon)d_n(T). \tag{2.7}$$

We put $M=N^\perp$ and fix $x\in B_X$ and $y\in N$ according to (2.7) and obtain

$$\sup_{\substack{\varphi\in N^\perp\\ ||\varphi||_{Y'}\le 1}} \sup_{x\in B_X} |\varphi(Tx)| = \sup_{\substack{\varphi\in N^\perp\\ ||\varphi||_{Y'}\le 1}} \sup_{x\in B_X} |\varphi(Tx-y)| \le \sup_{\substack{\varphi\in N^\perp\\ ||\varphi||_{Y'}\le 1}} \sup_{x\in B_X} |||\varphi||_{Y'}\cdot ||Tx-y||_X \le (1+\varepsilon)d_n(T).$$

Hence $c_n(T')\le(1+\varepsilon)d_n(T)$ and we let $\varepsilon\to 0$ to finish the proof.

Step 3. Proof of (2.6).

This is the most complicated step. If X and Y would be reflexive, then the proof would be trivial, because then $d_n(T)=d_n(T'')\le c_n(T')$. If this is not the case, we use the *principle of local reflexivity.*

Lemma 36. *Let Y be a Banach space and let $M\subset Y''$ be a finite-dimensional subspace of Y''. Then for every $\varepsilon>0$ there exists $R\in\mathcal{L}(M,Y)$, such that $||R|\mathcal{L}(M,Y)||<1+\varepsilon$ and $R\varepsilon_Y y=y$ for all $y\in Y$ with $\varepsilon_Y y\in M$.*

We now come back to the proof of (2.6).

We already know, that $c_n(T')=d_n(T'')$. So, it is enough to prove, that $d_n(T'')\ge d_n(T)$.

Let $\varepsilon>0$. Then there is a subspace $N\subset Y''$ with $\dim N<n$, such that

$$\inf_{y\in N} ||T''x-y||_{Y''} < (1+\varepsilon)d_n(T''). \tag{2.8}$$

Let $\{x_1,\dots,x_k\}\subset X$ be such that $\{Tx_1,\dots,Tx_n\}$ is an ε-net for $T(B_X)$. Let $M\subset Y''$ be the span of N and $\{\varepsilon_Y(Tx_1),\dots,\varepsilon_Y(Tx_k)\}$ and let $R\in\mathcal{L}(M,Y)$ be the mapping from the principle of local reflexivity, i.e.

$$||R|\mathcal{L}(M,Y)||<1+\varepsilon \quad\text{and}\quad R\varepsilon_Y(Tx_i)=Tx_i,\quad i=1,\dots,k.$$

Finally, let $L=R(N)\subset Y$ be a subspace of Y with dimension smaller than n.

According to (2.8), we find $z_1,\dots,z_k\in N$, such that

$$||T''(\varepsilon_X x_i)-z_i||_{Y''}<d_n(T'').$$

Altogether, we get for all $i=1,\dots,k$

$$\begin{aligned}\inf_{y\in L}||Tx_i-y||_Y &= \inf_{y\in L}||R\varepsilon_Y(Tx_i)-y||_Y = \inf_{y\in L}||R(T''\varepsilon_X x_i)-y||_Y\\ &\le ||R(T''\varepsilon_X x_i)-R(z_i)||_Y \le ||R|\mathcal{L}(M,Y)||\cdot||T''\varepsilon_X x_i - z_i||_{Y''}\\ &\le (1+\varepsilon)^2 d_n(T'').\end{aligned}$$

If $x\in B_X$, we get for appropriate $i=1,\dots,k$

$$\inf_{y\in L}||Tx-y||_Y\le ||Tx-Tx_i||_Y+\inf_{y\in L}||Tx_i-y||_Y\le\varepsilon+(1+\varepsilon)^2d_n(T'')$$

and the result follows. □

Remark 16. The inequality (2.6) does not hold for arbitrary linear operators. The classical counterexample is the identity $id:\ell_1\to c_0$ with $id':\ell_1\to\ell_\infty$. It may be shown, that $d_n(id)=1$ and $c_n(id')=1/2$ for all $n=2,3,\dots$.

Theorem 37. *Let X and Y be two Banach spaces and let $T\in\mathcal{K}(X,Y)$. Then*

$$a_n(T)=a_n(T'),\quad n\in\mathbb{N}.$$

Proof. We prove only the easy part, namely $a_n(T')\le a_n(T)$. The proof of the reverse inequality uses again the principle of local reflexivity and resembles the proof of Theorem 35.

Let $P\in\mathcal{L}(X,Y)$ with rank $P<n$. We consider $P'\in\mathcal{L}(Y',X')$ defined, as usually, by

$$(P'\varphi)(x)=\varphi(Px),\quad x\in X,\quad \varphi\in Y'.$$

Then

$$\begin{aligned}||T'-P'|\mathcal{L}(Y',X')||&=\sup_{\varphi\in B_{Y'}}||(T'-P')(\varphi)||_{X'}=\sup_{\varphi\in B_{Y'}}\sup_{x\in B_X}|[(T'-P')(\varphi)](x)|\\&=\sup_{\varphi\in B_{Y'}}\sup_{x\in B_X}|\varphi((T'-P')(x))|\le\sup_{\varphi\in B_{Y'}}\sup_{x\in B_X}||\varphi||_{Y'}\cdot||(T-P)(x)||_Y\\&\le\sup_{x\in B_X}||(T-P)(x)||_Y=||T-P|\mathcal{L}(X,Y)||.\end{aligned}$$

Taking the infimum over all P's finishes the proof. □

2.4 Approximation, Gelfand and Kolmogorov numbers of $id:\ell_p^m\to\ell_q^m$

Theorem 38. *Let $0<p\le\infty$ and let $\sigma=(\sigma_1,\sigma_2,\dots)$ with $\sigma_1\ge\sigma_2\ge\dots\ge0$ be a non-increasing sequence. We define the diagonal operator D_σ as*

$$D_\sigma:\ell_p\to\ell_p,\quad D_\sigma x=(\sigma_1x_1,\sigma_2x_2,\dots).$$

Then

$$c_n(D_\sigma)=d_n(D_\sigma)=\sigma_n,\qquad n\in\mathbb{N}.$$

Proof. The proof follows the same pattern as the proof of Theorem 23. □

2.4.1 Extreme points and the Krein-Milman theorem

Definition 39. Let X be a vector space and let $K\subset X$ be convex. Then

a) $F\subset K$ is called *side*, if F is convex and it holds

$$x_1,x_2\in K,0<\lambda<1,\lambda x_1+(1-\lambda)x_2\in F\implies x_1,x_2\in F.$$

b) $x \in K$ is called *extreme point*, if $\{x\} \subset K$ is a side of K. In other words, if

$$x_1, x_2 \in K, 0 < \lambda < 1, \lambda x_1 + (1-\lambda)x_2 = x \implies x_1 = x_2 = x$$

holds.

The set of all extreme points of K is denoted as Ext K.

Theorem 40. *Let X be a Banach space.*[22] *Let $K \subset X$ be compact, convex and non-empty. Then*
a) Ext $K \neq \emptyset$,
b) $K = \overline{conv}$ Ext K.

Proof. Let $\mathcal{F}$ denotes the set of all closed, non-empty sides of K. Then the following holds:

- $\mathcal{F} \neq \emptyset$, while $K \in \mathcal{F}$,
- if $\{F_\alpha\}_{\alpha \in A} \subset \mathcal{F}$ is a subsystem of $\mathcal{F}$ with $F_\alpha \subset F_{\alpha'}$ or $F_\alpha \supset F_{\alpha'}$ for every $\alpha, \alpha' \in A$, then $\bigcap_{\alpha \in A} F_\alpha \in \mathcal{F}$,
- $\mathcal{F}$ is inductively ordered with respect to inclusion.

According to The Lemma of Zorn, there is a minimal element $F_0 \in \mathcal{F}$.

We show, that F_0 consists of only one point.

Let $x_0, y_0 \in F_0$ with $x_0 \neq y_0$. Then (Hahn-Banach!) there is an $x' \in X'$, such that

$$\text{Re } x'(x_0) < \text{Re } x'(y_0).$$

Then the set

$$F_1 = \{x \in F_0 : \text{Re } x'(x) = \sup_{y \in F_0} \text{Re } x'(y)\}$$

is non-empty (F_0 is compact and x' is continuous). Further is F_1 closed and a side of F_0 and hence also of K, i.e. $F_1 \in \mathcal{F}$. But $x_0 \notin F_1$ - which is a contradiction with minimality of F_0.

To prove b), put

$$K_1 = \overline{conv} \text{ Ext } K. \tag{2.9}$$

Then $K_1 \subset K$ is closed (and hence also compact) and (according to a) also non-empty. If $K \neq K_1$, then there are $x \in K \setminus K_1$, $\varepsilon > 0$ and (Hahn-Banach!) $x' \in X'$, such that

$$\text{Re } x'(x) \leq \text{Re } x'(x_0) - \varepsilon \quad \text{for all} \quad x \in K_1.$$

We consider

$$F = \{x \in K : \text{Re } x'(x) = \sup_{y \in K} \text{Re } x'(y)\}.$$

Then one may show as above, that F is a closed, non-empty side in K. According to a), Ext $F \neq \emptyset$. Because of Ext F $\subset$ Ext K, there is an $e \in$ Ext K with $e \notin K_1$, which is a contradiction with (2.9). $\square$

[22] The theorem holds also for locally convex Hausdorff spaces(with the same proof), which enables to apply the statement also to weak topologies.

Lemma 41. *Let N be a subspace of ℓ_∞^m with* codim $N < n$*. Then there exists $x = (x_1, \ldots, x_m) \in N$, such that $||x||_\infty = 1$ and $\#\{k : |x_k| = 1\} \geq m - n + 1$.*

Proof. Let x be an extreme point of $B_{\ell_\infty^m} \cap N$. Let

$$K := \{k : |x_k| = 1\} \quad \text{and} \quad M := \{y \in \ell_\infty^m : y_k = 0 \text{ for } k \in K\}.$$

Clearly $\#K + \dim M = m$. Suppose, that $\#K \leq m - n$. Then $\dim M \geq n$ and therefore $M \cap N \neq \{0\}$. Hence, there is an $y \in M \cap N$ with $||y||_\infty = 1$.

As

$$\delta := 1 - \max\{|x_k| : k \notin K\} > 0,$$

it follows, that $x \pm \delta y \in B_{\ell_\infty^m} \cap N$. Hence, x cannot be an extreme point, which is a contradiction. □

Lemma 42. *Let $0 < q < p \leq \infty$. If $|x_{n+1}| \leq \min(|x_1|, \ldots, |x_n|)$, then*

$$\frac{\left\{\sum_{j=1}^{n+1} |x_j|^q\right\}^{1/q}}{\left\{\sum_{j=1}^{n+1} |x_j|^p\right\}^{1/p}} \geq \frac{\left\{\sum_{j=1}^{n} |x_j|^q\right\}^{1/q}}{\left\{\sum_{j=1}^{n} |x_j|^p\right\}^{1/p}}.$$

Proof. Let

$$\alpha = \left\{\sum_{j=1}^{n} |x_j|^p\right\}^{1/p} \quad \text{and} \quad \beta = \left\{\sum_{j=1}^{n} |x_j|^q\right\}^{1/q}.$$

As $p > q$, we have

$$\left|\frac{x_j}{x_{n+1}}\right|^q \leq \left|\frac{x_j}{x_{n+1}}\right|^p \quad \text{for all} \quad j = 1, \ldots, n.$$

Summing over $j = 1, \ldots, n$ gives

$$\left(\frac{\beta}{|x_{n+1}|}\right)^q \leq \left(\frac{\alpha}{|x_{n+1}|}\right)^p.$$

and

$$\left(1 + \left(\frac{|x_{n+1}|}{\beta}\right)^q\right)^{p/q} \geq 1 + \left(\frac{|x_{n+1}|}{\beta}\right)^q \geq 1 + \left(\frac{|x_{n+1}|}{\alpha}\right)^p.$$

Finally, we get

$$\frac{\left\{\sum_{j=1}^{n+1} |x_j|^q\right\}^{1/q}}{\left\{\sum_{j=1}^{n+1} |x_j|^p\right\}^{1/p}} = \frac{(\beta^q + |x_{n+1}|^q)^{1/q}}{(\alpha^p + |x_{n+1}|^p)^{1/p}} = \frac{\beta(1 + |x_{n+1}|^q/\beta^q)^{1/q}}{\alpha(1 + |x_{n+1}|^p/\alpha^p)^{1/p}} \geq \frac{\beta}{\alpha}.$$

□

Theorem 43. *Let $0 < q \le p \le \infty$. Then*

$$a_n(id : \ell_p^m \to \ell_q^m) = c_n(id : \ell_p^m \to \ell_q^m) = (m-n+1)^{1/q-1/p}.$$

If $1 \le q \le p \le \infty$, then also

$$d_n(id : \ell_p^m \to \ell_q^m) = (m-n+1)^{1/q-1/p}.$$

Proof. For the estimate from above, we consider the operator

$$P_{n-1} : \ell_p^m \to \ell_q^m, \quad P_{n-1}(x_1,\dots,x_m) = (x_1,\dots,x_{n-1},0,\dots).$$

This shows, that

$$c_n(id : \ell_p^m \to \ell_q^m) \le a_n(id : \ell_p^m \to \ell_q^m) \le ||id - P_{n-1}|\mathcal{L}(\ell_p^m,\ell_q^m)|| = (m-n+1)^{1/q-1/p}.$$

The estimate from below is the tricky part. Let $M \subset \ell_p^m$ be a subspace with codim $M < n$. Then Lemma 41 implies the existence of $x = (x_1,\dots,x_m)$ with $||x||_\infty = 1$ and $\#\{k : |x_k| = 1\} \ge m-n+1$. Then we get by Lemma 42

$$||id : M \to \ell_q^m|| \ge \frac{||x||_q}{||x||_p} = \frac{\left(\sum_{j=1}^m |x_j|^q\right)^{1/q}}{\left(\sum_{j=1}^m |x_j|^p\right)^{1/p}} \ge \frac{\left(\sum_{j\in K} |x_j|^q\right)^{1/q}}{\left(\sum_{j\in K} |x_j|^p\right)^{1/p}} \ge (m-n+1)^{1/q-1/p},$$

hence

$$c_n(id : \ell_p^m \to \ell_q^m) \ge (m-n+1)^{1/q-1/p}.$$

If $p = \infty$, only notational changes are necessary.

The result for the Kolmogorov numbers then follows by duality. □

Definition 44. Let H_1 and H_2 be two (separable) complex Hilbert spaces and let $T \in \mathcal{L}(H_1,H_2)$. Further let $\{u_i\}_{i\in I}$ be an orthonormal basis of H_1. Then the *Hilbert-Schmidt norm* of T is defined as

$$||T|\mathfrak{S}|| = \left(\sum_{i\in I} ||Tu_i||_{H_2}^2\right)^{1/2}.$$

Remark 17. The Hilbert-Schmidt norm does not depend on the choice of the orthonormal basis $\{u_i\}_{i\in I}$, cf. Exercises 33.

Theorem 45. Stechkin 1954.

$$a_n(id : \ell_1^m \to \ell_2^m) = \left(\frac{m-n+1}{m}\right)^{1/2} \quad \textit{for } n = 1,\dots,m.$$

Proof. Step 1. Estimate from below.

Let

$$L \in \mathcal{L}(\ell_1^m,\ell_2^m) \quad \text{with} \quad \text{rank } L < n.$$

Let $P \in \mathcal{L}(\ell_2^m, \ell_2^m)$ be the orthogonal projection of ℓ_2^m onto the orthogonal complement of the range of L, i.e. kern P $= L(\ell_2^m)$.

Then $||P|\mathfrak{S}||^2 \geq m - n + 1$ and

$$\sum_{k=1}^{m} ||Pe_k|\ell_2^m||^2 \geq m - n + 1,$$

where $\{e_k\}_{k=1}^m$ are the canonical unit vectors of $\mathbb{R}^m$.

Finally,

$$||id - L|\mathcal{L}(\ell_1^m, \ell_2^m)|| \geq ||P(id - L)|\mathcal{L}(\ell_1^m, \ell_2^m)|| \geq \max\{||Pe_k|\ell_2^m|| : k = 1, \dots, m\} \geq \left(\frac{m-n+1}{m}\right)^{1/2}.$$

Step 2. Estimate from above.

According to the Theorem 29, we have

$$a_n(id : \ell_1^m \to \ell_2^m) = d_n(id : \ell_1^m \to \ell_2^m) = \inf\{||id - P \circ id|\mathcal{L}(\ell_1^m, \ell_2^m)|| : \text{rank } P < n\},$$

where the infimum runs over all projections with rank $P < n$.

Using the fact, that

$$||A|\mathcal{L}(\ell_1^m, \ell_2^m)|| = \max_{j=1,\dots,m} ||Ae_j||_2,$$

we may rewrite this as

$$a_n(id : \ell_1^m \to \ell_2^m) = \inf\{\max_{j=1,\dots,m} ||(id - P \circ id)e_j||_2 : \text{rank } P < n\}.$$

Finally, using orthogonality of P, we get

$$a_n(id : \ell_1^m \to \ell_2^m) = \inf\{\max_{j=1,\dots,m} (1 - ||Pe_j||_2^2)^{1/2} : \text{rank } P < n\}. \tag{2.10}$$

The proof is then easily finished with the help of

Lemma 46. *Let* $1 \leq n \leq m$ *and* $\{\pi_j\}_{j=1}^m$ *with*

$$\sum_{j=1}^{m} \pi_j^2 = n \quad \textit{and} \quad 0 \leq \pi_j \leq 1 \quad \textit{for} \quad j = 1, \dots, m.$$

Then there is an orthogonal projection $P : \ell_2^m \to \ell_2^m$ *with* rank $P = n$ *and*

$$||Pe_j||_2 = \pi_j, \quad \textit{for all} \quad j = 1, \dots, m.$$

It is enough to choose $\pi_j^2 = \frac{n-1}{m}$ and (2.10) becomes

$$a_n(id : \ell_1^m \to \ell_2^m) \leq \max_{j=1,\dots,m} (1 - ||Pe_j||_2^2)^{1/2} = \left(\frac{m-n+1}{m}\right)^{1/2}.$$

□

We return to the proof of Lemma 46.

Proof. The proof goes by induction over n. If $n = 1$, then we set

$$Py =< x, y > x,$$

where $x = (\pi_1, \pi_2, \dots, \pi_m)$. Then rank $P = 1$ and

$$||Pe_j||_2^2 = | < x, e_j > |^2 \cdot ||x||_2^2 = \pi_j^2, \quad j = 1, \dots, m.$$

We suppose, that the assertion has been proved for some n and consider the sequence

$$1 \geq \pi_1 \geq \dots \geq \pi_m \geq 0 \quad \text{with} \quad \sum_{j=1}^{m} \pi_j^2 = n + 1.$$

Then there exists a natural number k (which we shall fix for the rest of the proof), such that

$$\sum_{j=1}^{k} \pi_j^2 \leq 1 < \sum_{j=1}^{k+1} \pi_j^2.$$

We define

$$\sigma_i = \begin{cases} \pi_i, & i = 1, \dots, k-1 \\ \left(1 - \sum_{j=1}^{k-1} \pi_j^2\right)^{1/2}, & i = k, \\ \left(\sum_{j=1}^{k+1} \pi_j^2 - 1\right)^{1/2}, & i = k+1, \\ \pi_i, & i = k+2, \dots, m. \end{cases}$$

Then

$$\sum_{j=1}^{k} \sigma_j^2 = 1 \quad \text{and} \quad \sum_{j=k+1}^{m} \sigma_j^2 = \sum_{j=1}^{m} \pi_j^2 - 1 = n.$$

Hence, there is an orthogonal projection P with rank $P = n$ and

$$||Pe_j||_2^2 = 0 \text{ if } j = 1, \dots, k \quad \text{and} \quad ||Pe_j||_2^2 = \sigma_j^2 \text{ if } j = k+1, \dots, m.$$

We define

$$x_0 = (\sigma_1, \dots, \sigma_k, 0, \dots, 0) \in \mathbb{R}^m \quad \text{and} \quad P_0 y :=< x_0, y > x_0 + Py.$$

Then

- $||x_0||_2 = 1$,
- $Px_0 = 0$,
- P_0 is an orthogonal projection,
- rank $P_0 = n + 1$,
- $||P_0 e_i||_2^2 = \sigma_i^2$ for $i = 1, 2, \dots, m$.

For $0 \le \alpha \le 1$, we define the orthonormal sequence $\{u_i\}_{i=1}^m$ by

$$u_i^\alpha = \begin{cases} e_i, & i = 1, \dots, k-1 \\ (1-\alpha^2)^{1/2} e_k + \alpha e_{k+1}, & i = k, \\ -\alpha e_k + (1-\alpha^2)^{1/2} e_{k+1}, & i = k+1, \\ e_i, & i = k+2, \dots, m. \end{cases}$$

Hence,

$$u_i^0 = \begin{cases} e_i, & i = 1, \dots, k-1 \\ e_k, & i = k, \\ e_{k+1}, & i = k+1, \\ e_i, & i = k+2, \dots, m, \end{cases} \quad \text{and} \quad u_i^1 = \begin{cases} e_i, & i = 1, \dots, k-1 \\ e_{k+1}, & i = k, \\ -e_k, & i = k+1, \\ e_i, & i = k+2, \dots, m \end{cases}$$

and u_i^α represents a continuous way between these two extrema.
As

$$\begin{pmatrix} (1-\alpha^2)^{1/2} & \alpha \\ -\alpha & (1-\alpha^2)^{1/2} \end{pmatrix} \cdot \begin{pmatrix} (1-\alpha^2)^{1/2} & -\alpha \\ \alpha & (1-\alpha^2)^{1/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\begin{pmatrix} (1-\alpha^2)^{1/2} & -\alpha \\ \alpha & (1-\alpha^2)^{1/2} \end{pmatrix} \cdot \begin{pmatrix} (1-\alpha^2)^{1/2} & \alpha \\ -\alpha & (1-\alpha^2)^{1/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

we see, that the operator

$$U_\alpha y := \sum_{i=1}^m < e_i, y > u_i^\alpha$$

is unitary. It follows, that if $P_\alpha = U_\alpha^* P_0 U_\alpha$,[23] then

$$\begin{aligned} < x - P_\alpha x, P_\alpha y > &=< x - U_\alpha^* P_0 U_\alpha x, U_\alpha^* P_0 U_\alpha y >=< U_\alpha x - U_\alpha U_\alpha^* P_0 U_\alpha x, U_\alpha U_\alpha^* P_0 U_\alpha y > \\ &=< U_\alpha x - P_0 U_\alpha x, P_0 U_\alpha y >= 0. \end{aligned}$$

Hence, P_α is an orthogonal projection for every $0 \le \alpha \le 1$. Furthermore, we have

- $||P_\alpha e_i||_2^2 = \sigma_i^2 = \pi_i^2$ for all $i = 1, \dots, k-1$ and $i = k+2, \dots, m$ and all $0 \le \alpha \le 1$,
- $||P_0 e_k||_2^2 = \sigma_k^2$,
- $||P_1 e_k||_2^2 = \sigma_{k+1}^2$,
- $\sigma_k^2 \ge \pi_k^2 \ge \pi_{k+1}^2 \ge \sigma_{k+1}^2$.

Since $||P_\alpha e_k||_2^2$ depends continuously on α, there is an $\tilde\alpha \in [0, 1]$ with $||P_{\tilde\alpha} e_k||_2^2 = \pi_k^2$, which implies also $||P_{\tilde\alpha} e_{k+1}||_2^2 = \pi_{k+1}^2$. Hence $P_{\tilde\alpha}$ is the desired projection. □

Theorem 47.

$$a_n(id : \ell_1 \to c_0) = 1 \quad \text{for} \quad n = 1, 2, \dots.$$

[23] Note, that P_0 according to this new definition coincides with the P_0 defined above.

Proof. The estimate from above is trivial. For the estimate from below, we consider an arbitrary $A \in \mathcal{F}(\ell_1, c_0)$. Then A is compact and

$$A(B_{\ell_1}) \subset \bigcup_{j=1}^{N} y^j + \varepsilon B_{c_0}$$

for every $\varepsilon > 0$ and suitable $N \in \mathbb{N}$ and $y^1, \dots, y^N \in c_0$.

We choose an $n \in \mathbb{N}$, so that $|y_n^j| < \varepsilon$ for all $j = 1, \dots, N$. Then for some j' we have $||Ae_n - y^{j'}||_\infty \le \varepsilon$ and hence

$$||id - A|\mathcal{L}(\ell_1, c_0)|| \ge ||(id - A)e_n||_\infty \ge |1 - (Ae_n)_n| \ge |1 - y_n^{j'}| - |y_n^{j'} - (Ae_n)_n| \ge 1 - 2\varepsilon.$$

This proves, that

$$||id - A|\mathcal{L}(\ell_1, c_0)|| \ge 1.$$

□

Theorem 48.

$$a_n(id : \ell_1 \to \ell_\infty) = 1/2 \quad for \quad n = 2, \dots.$$

Proof. Let

$$A_0 y := \frac{x}{2} \cdot \sum_{i=1}^{\infty} x_i y_i,$$

where $x = (1, 1, 1, \dots)$. Then rank $A_0 = 1$ and

$$||id - A|\mathcal{L}(\ell_1, \ell_\infty)|| = \frac{1}{2}.$$

To prove the estimate from below, we suppose, that there is an $A \in \mathcal{F}(\ell_1, \ell_\infty)$, such that

$$||id - A|\mathcal{L}(\ell_1, \ell_\infty)|| = \sup_{j,k \in \mathbb{N}} |(e_j - Ae_j)_k| < 1/2.$$

But the estimate

$$\begin{aligned} ||Ae_j - Ae_k||_\infty &\ge |(Ae_j)_k - (Ae_k)_k| = |1 - (Ae_j)_k - (1 - (Ae_k)_k)| \ge 1 - |(Ae_j)_k| - |1 - (Ae_k)_k| \\ &\ge 1 - 2||id - A|\mathcal{L}(\ell_1, \ell_\infty)|| > 0. \end{aligned}$$

Hence $A(B_{\ell_1})$ is not precompact in ℓ_∞, which is a contradiction. □

Next, we consider the approximation numbers of $id : \ell_1^m \to \ell_\infty^m$. We start with the following simple observation

$$||id - A|\mathcal{L}(\ell_1^m, \ell_\infty^m)|| = \max_{j=1,\dots,m} ||(id - A)e_j||_\infty = \max_{j,k=1,\dots,m} |e_{j,k} - (Ae_j)_k| = \max_{j,k=1,\dots,m} |\delta_k^j - A_{k,j}|.$$

The last expression is to be minimalised through all matrices $A = (A_{i,j})_{i,j=1}^m$ with rank $A < n$.

The following lemma describes an easy way, how to produce a (symmetric) matrix A with small rank.

Lemma 49. *Let $1 \le n \le m$ and let $x_1, \dots, x_m \in \mathbb{R}^n$. Then the matrix*

$$A = (A_{i,j})_{i,j=1}^m \quad with \quad A_{i,j} = \langle x_i, x_j \rangle$$

has rank $A \le n$.

Proof. Let $y = (y_1, \dots, y_m) \in \mathbb{R}^m$ and write

$$\begin{aligned} Ay &= \sum_{i=1}^m \sum_{j=1}^m < x_i, x_j > y_j e_i = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^n x_{i,k} x_{j,k} < y, e_j > e_i = \sum_{k=1}^n < y, \sum_{j=1}^m x_{j,k} e_j > \sum_{i=1}^m x_{i,k} e_i \\ &= \sum_{k=1}^n < y, \chi_k > \chi_k, \end{aligned}$$

where $\chi_k = \sum_{j=1}^m x_{j,k} e_j \in \mathbb{R}^m$. □

So, for the estimate of $a_n(id : \ell_1^m \to \ell_\infty^m)$ we need to construct m vectors $x_1, \dots, x_m$ in $\mathbb{R}^n$, such that

- $||x_i||_2 = 1$, i.e. $(x_i, x_i) = 1$ for all $i = 1, \dots, m$,
- $| < x_i, x_j > |$ is as small as possible for $i \neq j$.

We shall give two such constructions. One probabilistic, based on Khintchine's inequalities and another explicit, based on polynmials on $GF(p)$.

Lemma 50. Khintchine's Inequality *Let*

$$r_n(t) = \text{sign } \sin(2^n \pi t), \quad n \in \mathbb{N}, \quad t \in [0, 1].$$

Then for every $1 \le p < \infty$, there are two constants A_p and B_p, such that

$$A_p \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2} \le \left(\int_0^1 \left| \sum_{n=1}^m a_n r_n(t) \right|^p dt \right)^{1/p} \le B_p \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2}$$

for every $m \in \mathbb{N}$ and all $a_1, \dots, a_m \in \mathbb{R}$.

Proof. See [HHZ, pp. 205-206]. □

Remark 18. a) Let us remark (without proof), that $B_p \le ([p/2] + 1)^{1/2}$.
b) The Khintchine's inequality may be rewritten using sums. Then it reads

$$A_p \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2} \le \left(\frac{1}{2^m} \sum_{e \in \{-1,+1\}^m} |< a, e >|^p \, dt \right)^{1/p} \le B_p \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2}$$

for every $m \in \mathbb{N}$ and every $a \in \mathbb{R}^m$.

Lemma 51. *There are $x_1, \dots, x_m \in \ell_2^n$ such that $||x_i||_2 = 1$ for all $i = 1, 2, \dots, m$ and*

$$|(x_i, x_j)| \le 2 \left[\frac{\log_2 m}{n} \right]^{1/2} \quad \text{for} \quad i \neq j.$$

Proof. If $m \le n$, then we may take an orthonormal family $(x_1, \dots, x_m) \subset \mathbb{R}^n$. Let us suppose, that $m \ge n$ and that the result has already been proved for this m. Then

$$\sum_{i=1}^{m} \sum_{e \in \{-1,+1\}^n} | < x_i, e > |^p \le B_p^p m 2^n.$$

Hence, at least for one $e \in \{-1, +1\}^n$,

$$\sum_{i=1}^{m} | < x_i, e > |^p \le B_p^p m.$$

We put $x_{m+1} := n^{-1/2} e$. Hence $||x_{m+1}||_2 = 1$ and

$$| < x_i, x_{m+1} > | \le B_p m^{1/p} n^{-1/2} \quad \text{for} \quad i = 1, 2, \dots, m.$$

By choosing $p := \log_2 m$, we obtain

$$| < x_i, x_{m+1} > | \le 2 \left[\frac{\log_2 m}{n} \right]^{1/2} \quad \text{for} \quad i = 1, 2, \dots, m.$$

□

Lemma 52. *Let $0 < \lambda < 1$. Then there is a constant $c_\lambda > 0$, such that for every $m^\lambda \le n \le m$, there are m unit vectors $x_1, x_2, \dots, x_m$ in ℓ_2^n, such that*

$$|(x_i, x_j)| \le \frac{c_\lambda}{n^{1/2}}, \qquad i \ne j.$$

Proof. Let p be a prime number and let $GF(p)$ be the *Galois field*, i.e. the set $\{0, 1, \dots, p-1\}$ equipped with addition and multiplication modulo p. Let $k \in \mathbb{N}$ and let $\mathcal{P}_k$ denotes all the polynomes over $GF(p)$ with degree smaller or equal k. For every $\pi \in P_k$, we define a vector

$$x^\pi \in \ell_2^{p^2}, \qquad x^\pi_{i,j} = \begin{cases} 1 & \text{if } \pi(i) = j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence

- $||x^\pi||_2 = \sqrt{p}$ for every $\pi \in \mathcal{P}_k$,
- $|(x^\pi, x^\sigma)| \le k$ for all $\pi, \sigma \in \mathcal{P}_k$ with $\pi \ne \sigma$,
- one obtains $\#\mathcal{P}_k = p^{k+1}$ vectors.

Hence the system

$$\left\{ \frac{x^\pi}{\sqrt{p}} \right\}_{\pi \in \mathcal{P}_k}$$

has all the desired properties for

$$n = p^2, \quad \lambda = \frac{2}{k+1}, \quad m = p^{k+1}, \quad c_\lambda = k.$$

□

Using the vectors of Lemma 51 and Lemma 52, respectively, one proves

Theorem 53. *a)*

$$a_n(id : \ell_1^m \to \ell_\infty^m) \leq 3 \left[\frac{\log_2(m+1)}{n} \right]^{1/2}, \quad n = 1, \dots, m.$$

b) Let $0 < \lambda < 1$*. Then there is a* $c_\lambda > 0$*, such that*

$$a_n(id : \ell_1^m \to \ell_\infty^m) \leq \frac{c_\lambda}{n^{1/2}}, \quad m^\lambda \leq n \leq m.$$

Proof. Step 1. Proof of a)

Let $x_1, \dots, x_n$ be the vectors constructed in Lemma 51 and define $A = \{A_{i,j}\}_{i,j=1}^m = \{< x_i, x_j >\}_{i,j=1}^m$. Then

$$a_{n+1}(id : \ell_1^m \to \ell_\infty^m) \leq ||id - A|\mathcal{L}(\ell_1^m, \ell_\infty^m)|| \leq 2 \left[\frac{\log_2 m}{n} \right]^{1/2} \leq 3 \left[\frac{\log_2(m+1)}{n+1} \right]^{1/2}.$$

Step 2. Proof of b) follows in the same way. □

References

[1] D. E. Edmunds and H. Triebel, *Function spaces, entropy numbers, differential operators*, Cambridge University Press, 1996.

[2] D. D. Haroske and H. Triebel, *Distributions, Sobolev spaces, elliptic equations*, EMS Publishing House, 2008.

[3] H. Heuser, *Lehrbuch der Analysis*, Teil 1, 6. Auflage, B. G. Teubner Stuttgart 1989.

[4] A. Pietsch, *Operator ideals*, Deutsch. Verlag Wiss., Berlin, 1978; North-Holland, Amsterdam-London-New York-Tokyo, 1980.

[5] A. Pietsch, *Eigenvalues and s-numbers*, Cambridge University Press, 1987, Cambridge.