Blueprint:
s-Numbers of Bounded Linear Operators between Banach Spaces

Mario Ullrich

September 13, 2026

Contents

Chapter 1

Setting and prerequisites

Throughout the blueprint, $\mathbb{K}$ denotes a non-trivially normed field — in practice $\mathbb{R}$ or $\mathbb{C}$ — and X, Y, Z, W are Banach spaces over $\mathbb{K}$. We write $\mathcal{L}(X, Y)$ for the space of bounded linear operators $X \to Y$ with the operator norm, and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$; all sequences are indexed starting from 0.

Each statement is accompanied by the name of the Lean declaration that formalises it. Every statement and every proof here is formalised: the project compiles without `sorry`, so each result is verified by the Lean kernel.

Definition 1.1 (Rank of a continuous linear map)**.** For $S \in \mathcal{L}(X, Y)$ the *rank* of S is the cardinal dimension of its range,

$$\operatorname{rank}(S) := \dim_{\mathbb{K}} \operatorname{range}(S) \in \mathrm{Card}.$$

[Lean: `ContinuousLinearMap.rank`]

Chapter 2

Axioms of an s-number sequence

Definition 2.1 (s-number sequence; Pietsch axioms)**.** A family

$$s \;=\; (s_n)_{n\in\mathbb{N}_0}, \qquad s_n : \mathcal{L}(X,Y) \longrightarrow [0,\infty),$$

defined for all pairs of Banach spaces (X,Y) over $\mathbb{K}$ is called an s-*number sequence* if:

(S1) $\|S\| = s_0(S) \geq s_1(S) \geq \cdots \geq 0$ for every S.

(S2) $s_n(S+T) \leq s_n(S) + \|T\|$.

(S3) $s_n(BSA) \leq \|B\|\, s_n(S)\, \|A\|$ for composable bounded operators A, B.

(S4) $s_n(S) = 0$ whenever $\operatorname{rank}(S) \leq n$.

(S5) $s_n(\mathrm{id}_{\ell_2^{n+1}}) = 1$.

[Lean: `SNumbers.IsSNumberSequence`]

Definition 2.2 (Strict s-number sequence)**.** An s-number sequence is *strict* if it additionally satisfies the strengthening

(S5′) $s_n(\mathrm{id}_X) = 1$ for every Banach space X with $\dim X > n$, not merely on $X = \ell_2^{n+1}$.

[Lean: `SNumbers.IsStrictSNumberSequence`]

Chapter 3

The five canonical examples

We construct the five classical s-number sequences and verify the axioms (S1)–(S5) — and, where it holds, the strengthening (S5′) — for each of them.

3.1 Approximation numbers a_n

Definition 3.1 (Approximation number)**.** For $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,

$$a_n(S) := \inf\big\{\|S - A\| : A \in \mathcal{L}(X, Y), \operatorname{rank}(A) \le n\big\}.$$

[Lean: `SNumbers.approximationNumber`]

Theorem 3.2. *Let X, Y be Banach spaces over $\mathbb{K}$. Then the approximation numbers $a = (a_n)_n$ form a* strict *s-number sequence: they satisfy the axioms (S1)–(S5) and the strengthening (S5′).*
[Lean: `SNumbers.isStrictSNumberSequence_approximationNumber`]

Proof. Axioms (S1)–(S4) follow directly from the definition of a_n as an infimum: monotonicity and nonnegativity are inherited from the underlying set of approximants, subadditivity (S2) uses that A approximates $S + T$ whenever it approximates S up to $\|T\|$, the ideal property (S3) uses $\operatorname{rank}(BA'A) \le \operatorname{rank}(A')$, and (S4) is witnessed by the approximant $A = S$. For (S5′) let L have $\operatorname{rank}(L) \le n < \dim X$; then $\operatorname{range}(L)$ is a proper finite-dimensional, hence closed, subspace of X, and Riesz's lemma provides, for every $r < 1$, a unit vector x_0 with $\|x_0 - Lx_0\| \ge r$, so $\|\mathrm{id} - L\| \ge r$; letting $r \to 1$ and combining with $a_n(\mathrm{id}) \le \|\mathrm{id}\| = 1$ gives $a_n(\mathrm{id}_X) = 1$. The classical (S5) on ℓ_2^{n+1} is the special case $X = \ell_2^{n+1}$. The argument needs neither the SVD nor Auerbach's lemma. □

3.2 Bernstein numbers b_n

Definition 3.3 (Bernstein number)**.** For $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,

$$b_n(S) := \sup_{\dim M = n+1} \inf_{\substack{x \in M \\ x \neq 0}} \frac{\|Sx\|}{\|x\|}.$$

[Lean: `SNumbers.bernsteinNumber`]

Theorem 3.4. *Let $\mathbb{K}$ be a non-trivially normed field and X, Y normed spaces over $\mathbb{K}$. Then the Bernstein numbers $b = (b_n)_n$ form a* strict *s-number sequence: they satisfy (S1)–(S5) and the strengthening (S5′).*

[Lean: `SNumbers.isStrictSNumberSequence_bernsteinNumber`]

Proof. Write $\gamma(S, M) := \inf_{0 \neq x \in M} \|Sx\|/\|x\|$ for the gain of S on a subspace M, so that $b_n(S) = \sup_{\dim M = n+1} \gamma(S, M)$. On a one-dimensional $M = \operatorname{span}\{v\}$ the gain is exactly $\|Sv\|/\|v\|$, and the supremum over v gives $b_0(S) = \|S\|$; monotonicity in n holds because every $(n+2)$-dimensional subspace contains an $(n+1)$-dimensional one, and shrinking M only enlarges the gain. Subadditivity (S2) is the pointwise triangle inequality $\|(S+T)x\|/\|x\| \leq \|Sx\|/\|x\| + \|T\|$, pushed through the infimum and the supremum. For the ideal property (S3) fix an $(n+1)$-dimensional $M \subseteq W$: if A is not injective on M, the gain of BSA on M vanishes; if it is, then $M' := A(M)$ has dimension $n+1$ and $\gamma(BSA, M) \leq \|B\| \, \gamma(S, M') \, \|A\| \leq \|B\| \, b_n(S) \, \|A\|$. For the rank axiom (S4), an operator of rank $\leq n$ has nontrivial kernel on every $(n+1)$-dimensional subspace by rank–nullity, forcing the gain to zero. Finally (S5′): the identity has gain 1 on every nontrivial subspace, and $\dim X > n$ guarantees an $(n+1)$-dimensional subspace exists. □

The argument needs only a non-trivially normed scalar field $\mathbb{K}$ — no completeness, and no Riesz lemma.

3.3 Gelfand numbers c_n

Definition 3.5 (Gelfand number)**.** For $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,

$$c_n(S) := \inf\big\{\|S\restriction_M\| : M \subseteq X \text{ closed subspace}, \ \operatorname{codim} M \leq n\big\}.$$

[Lean: `SNumbers.gelfandNumber`]

Theorem 3.6. *Let $\mathbb{K}$ be a non-trivially normed field and X, Y normed spaces over $\mathbb{K}$. Then the Gelfand numbers $c = (c_n)_n$ form a* strict *s-number sequence: they satisfy (S1)–(S5) and the strengthening (S5′).*
[Lean: `SNumbers.isStrictSNumberSequence_gelfandNumber`]

Proof. The development is dual to the Kolmogorov one (Theorem 3.8) under the domain $\leftrightarrow$ codomain swap: where the Kolmogorov numbers push S down a quotient projection $\pi_V : Y \to Y/V$ for a small $V \subseteq Y$, the Gelfand numbers restrict S along the inclusion $\iota_M : M \hookrightarrow X$ for a subspace $M \subseteq X$ of small codimension. With $M = X$ one gets $c_0(S) = \|S\|$; (S2) is the triangle inequality applied to $(S + T) \circ \iota_M$. For the ideal property (S3), following [8, §2.4], take admissible M for S and pass to $M' := A^{-1}(M)$: it is closed, its codimension is at most that of M by the first isomorphism theorem applied to $\pi_M \circ A$, and $\|BSA(w)\| \leq \|B\| \, \|S \circ \iota_M\| \, \|A\| \, \|w\|$ for $w \in M'$. For (S4) take $M := \ker S$, which is closed with $\operatorname{codim} M = \operatorname{rank} S \leq n$ and $S \circ \iota_M = 0$. For (S5′), the inclusion ι_M is an isometry, so the identity restricted to any nontrivial M has norm 1, and $\operatorname{codim} M \leq n < \dim X$ forces $M \neq \{0\}$. Formalised in `Gelfand.lean`. □

The argument needs only a non-trivially normed scalar field $\mathbb{K}$; unlike its Kolmogorov dual (Theorem 3.8) it requires neither completeness of $\mathbb{K}$ nor Riesz's lemma.

3.4 Kolmogorov numbers d_n

Definition 3.7 (Kolmogorov number)**.** For $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,

$$d_n(S) := \inf_{\substack{V \subseteq Y \\ \dim V \leq n}} \sup_{\|x\| \leq 1} \inf_{v \in V} \|Sx - v\|.$$

[Lean: `SNumbers.kolmogorovNumber`]

Theorem 3.8. *Let $\mathbb{K}$ be a non-trivially normed field and X, Y normed spaces over $\mathbb{K}$. Then the Kolmogorov numbers $d = (d_n)_n$ form a* strict *s-number sequence: they satisfy (S1)–(S5) and the strengthening (S5′).*
[Lean: `SNumbers.isStrictSNumberSequence_kolmogorovNumber`]

Proof. The proof works with the operator-norm-on-quotient identity $d_n(S) = \inf_V \|\pi_V \circ S\|$, where $\pi_V : Y \to Y/V$ is the canonical projection onto the quotient. (S2) is the triangle inequality applied to $\pi_V \circ S$; for

(S3), an admissible V for S transports to the admissible subspace $B(V)$ for BSA, with the deviation bounded per point; for (S4) take $V = \operatorname{range} S$. For (S5′), given a subspace V of dimension $\leq n < \dim X$, its closure is proper and Riesz's lemma produces unit vectors almost at distance 1 from V, so $\|\pi_V\| = 1$ and $d_n(\mathrm{id}_X) = 1$. □

The (S5′) step additionally needs $\mathbb{K}$ *complete* (to close finite-dimensional subspaces via Riesz's lemma); it requires neither a densely normed $\mathbb{K}$ nor completeness of X.

3.4.1 Alternative: Pietsch's lifting identity

A second formalization of the Kolmogorov numbers, restricted to the Banach-space setting, lives in `KolmogorovLifting.lean`. It uses Pietsch's identity

$$d_n(S) = a_n(S \circ Q_X),$$

where a_n is the approximation number and $Q_X : \ell^1(B_X) \twoheadrightarrow X$ is the canonical summation surjection from the ℓ^1 space indexed by the closed unit ball B_X, given by

$$Q_X(\alpha) := \sum_{x \in B_X} \alpha(x) \cdot x.$$

Theorem 3.9. *For every Banach space X and every $S \in \mathcal{L}(X, Y)$, $d_n(S) = a_n(S \circ Q_X)$, and $d = (d_n)_n$ so defined is a* strict *s-number sequence: it satisfies (S1)–(S5) and (S5′).*

[Lean: `SNumbers.Lifting.kolmogorovNumber_strict` `SNumbers.Lifting.kolmogorovNumber_comp_comp_le`]

Proof. In this variant $d_n(S)$ is *defined* as $a_n(S \circ Q_X)$, so each axiom reduces to the corresponding fact for the approximation numbers (Theorem 3.2). Only the ideal property (S3) needs an extra construction: it factors as $(B \circ S \circ A) \circ Q_W = B \circ (S \circ Q_X) \circ T$ through a lift $T : \ell^1(B_W) \to \ell^1(B_X)$ that sends the basis vector δ_w to $c \cdot \delta_{c^{-1}Aw}$ for a scalar c with $\|A\| < |c|$, so that $\|T\| \leq |c|$; letting $|c| \searrow \|A\|$, using that the norm values of $\mathbb{K}$ are dense, yields the bound $\|B\| \, d_n(S) \, \|A\|$. The identification of this d_n with the canonical Kolmogorov number of Definition 3.7 is Pietsch's identity (Theorem 3.10). □

The lifting development assumes that $\mathbb{K}$ is a densely normed field (for the rescaling $|c| \searrow \|A\|$ in the ideal property (S3)) and that the source spaces X, W are complete (so that $Q_X(\alpha)$ converges as a series in X), in addition to completeness of $\mathbb{K}$ for (S5′). On Banach spaces the two formalisations agree, by Pietsch's identity (Theorem 3.10).

Theorem 3.10 (Pietsch's identity)**.** *For every Banach space X and every $S \in \mathcal{L}(X, Y)$, the canonical Kolmogorov number equals the approximation number of the lift: $d_n(S) = a_n(S \circ Q_X)$.*

[Lean: `SNumbers.kolmogorovNumber_eq_approx`]

Proof. The inequality $d_n(S) \le a_n(S \circ Q_X)$: for a rank-$\le n$ approximant L of $S \circ Q_X$, take $V := \operatorname{range} L$; since every unit vector x is $Q_X(\delta_x)$ and $\pi_V \circ L = 0$, one gets $\|\pi_V \circ S\| \le \|S \circ Q_X - L\|$. The reverse inequality: for an admissible V and $\varepsilon > 0$, pick $w_x \in V$ with $\|Sx - w_x\| < \|\pi_V \circ S\| + \varepsilon$ and lift to $L(\alpha) := \sum_x \alpha(x)\, w_x$ (a rank-$\le n$ operator into the finite-dimensional V), so that $\|S \circ Q_X - L\| \le \|\pi_V \circ S\| + \varepsilon$. $\square$

3.5 Hilbert numbers h_n

For the Hilbert numbers we work over $\mathbb{R}$ and use $L_2 := \ell_2(\mathbb{N}; \mathbb{R})$.

Definition 3.11 (Hilbert number)**.** For $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,

$$h_n(S) := \sup\left\{ \frac{a_n(B\, S\, A)}{\|B\|\, \|A\|} : A : L_2 \to X,\ B : Y \to L_2,\ A \neq 0,\ B \neq 0 \right\}.$$

[Lean: `SNumbers.hilbertNumber`]

Proposition 3.12. *Let X, Y be Banach spaces. Then the Hilbert numbers $h = (h_n)_n$ form an s-number sequence: they satisfy the axioms (S1)–(S5). The sequence h is not strict: (S5′) fails in general.*

[Lean: `SNumbers.isSNumberSequence_hilbertNumber`]

Proof. Each axiom is obtained by pushing the corresponding property of the approximation numbers a_n through the supremum defining h_n: nonnegativity, antitonicity, subadditivity (S2), the ideal property (S3) and the rank axiom (S4) transfer directly. For the normalisations, $h_0(S) = \|S\|$ combines the upper bound $h_n(S) \le \|S\|$ with a rank-one factorisation of arbitrary unit vectors through L_2, and $h_n(\mathrm{id}_{\ell_2^{n+1}}) = 1$ (S5) tests the supremum with the canonical embedding and projection between ℓ_2^{n+1} and L_2. $\square$

Chapter 4

The smallest and the largest s-numbers

Theorem 4.1. *On a Hilbert space, every s-number sequence agrees with the approximation numbers: $s_n(S) = a_n(S)$ for every continuous linear map S between Hilbert spaces.*

[Lean: `SNumbers.allSNumbers_eq_on_HilbertSpace`]

Proof. The upper bound is Theorem 4.2. For the lower bound $a_n(S) \leq s_n(S)$: given $c < a_n(S)$, the scalar factorisation (Theorem 6.8) gives contractions with $B \circ S \circ A = c \cdot \mathrm{id}_{\ell_2^{n+1}}$; chaining (S3) with the (S5) normalisation $s_n(\mathrm{id}_{\ell_2^{n+1}}) = 1$ and homogeneity yields $c \leq s_n(S)$. Letting $c \to a_n(S)$ gives $a_n(S) \leq s_n(S)$. □

Theorem 4.2 (Sandwich theorem)**.** *For every s-number sequence s and every operator S,*

$$h_n(S) \ \leq \ s_n(S) \ \leq \ a_n(S).$$

[Lean: `SNumbers.hilbertNumber_le_sn_le_approximationNumber`]

Proof. For the upper bound $s_n(S) \leq a_n(S)$, let A be any operator with $\mathrm{rank}(A) \leq n$; then axiom (S2) yields $s_n(S) \leq s_n(A) + \|S - A\| = \|S - A\|$ by (S4), and taking the infimum over A finishes. For the lower bound $h_n(S) \leq s_n(S)$: each ratio $a_n(BSA)/(\|B\|\|A\|)$ defining $h_n(S)$ has BSA an operator between Hilbert spaces, so the coincidence theorem (Theorem 4.1) gives $a_n(BSA) = s_n(BSA) \leq \|B\| \, s_n(S) \, \|A\|$ by (S3); take the supremum. □

Chapter 5

Basic auxiliary results

The companion library `BasicResults` (sibling of `SNumbers`) collects generic functional-analysis results that are independent of the s-numbers framework: Auerbach's lemma (this chapter), the singular value decomposition of compact operators (Chapter 6), determinant identities and the bordered-determinant formula, John's ellipsoid with the Kadets–Snobar and Garling–Gordon projection theorems (Section 7.1), the little Grothendieck bounds (Section 8.3.1), and the spectral projection underlying the Hilbert-space uniqueness theorem.

Theorem 5.1 (Auerbach's lemma)**.** *Every finite-dimensional normed space over $\mathbb{R}$ of positive dimension admits a basis (e_i) with $\|e_i\| = 1$ whose dual coordinate functionals also satisfy $\|e_i^*\| = 1$.*
[Lean: `exists_isAuerbachBasis`]

Proof. The strategy maximises $|\det|$ on a product of unit balls, then reads off the basis and dual functionals from the maximiser. □

Chapter 6

Singular value decomposition

The singular value decomposition (Schmidt representation) of a compact operator between Hilbert spaces is formalised in `BasicResults.SVD`, via the *singular-value iteration* (repeatedly peeling off the top singular pair). The auxiliary class of *approximable* operators (those whose approximation numbers a_n tend to zero) and its identification with the compact operators on Hilbert spaces live in the separate `AddOns` library; this split avoids relying on any approximation-property hypothesis on the underlying space.

6.1 Approximable operators

Definition 6.1 (Approximable operator)**.** $S \in \mathcal{L}(X, Y)$ is *approximable* if $a_n(S) \xrightarrow{n\to\infty} 0$, equivalently if S is the operator-norm limit of a sequence of finite-rank operators.

[Lean: `SVD.IsApproximable`]

Proposition 6.2. *The approximable operators form a closed two-sided operator ideal: a subspace of $\mathcal{L}(X, Y)$, closed under pre/post-composition with bounded operators, closed in operator-norm topology, and containing all finite-rank operators.*

[Lean: `SVD.isApproximable_iff_existsLimit SVD.isClosed_isApproximable`]

Proof. The equivalence with being a norm-limit of finite-rank operators follows by choosing near-optimal rank-$\leq n$ approximants L_n with $\|S - L_n\| < a_n(S) + \frac{1}{n+1}$, and conversely by squeezing $a_n(S) \leq \|S - L_n\|$. Stability under addition combines rank-$\leq \lfloor n/2 \rfloor$ and rank-$\leq \lceil n/2 \rceil$ approximants of the two summands, using that ranks are subadditive; stability under composition and closedness in operator norm are the axioms (S3) and (S2) of a_n passed through the limit. $\square$

Theorem 6.3. *Every approximable operator $S : X \to Y$ from a normed space X to a Banach space Y over a locally compact field $\mathbb{K}$ is compact. In particular every operator of finite rank is compact, since finite-rank operators are approximable (Proposition 6.2).*
[Lean: `SVD.IsApproximable.isCompactOperator SVD.isCompactOperator_of_rank_le`]

Proof. Each finite-rank approximant L_n corestricts to its range, a finite-dimensional subspace of Y, which is locally compact because $\mathbb{K}$ is; an operator into a locally compact space is compact, and post-composition with the inclusion preserves compactness. Since compactness passes to operator-norm limits, $S = \lim_n L_n$ is compact. $\square$

6.2 Schmidt representation

For the rest of this chapter, H_1, H_2 are Hilbert spaces over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Theorem 6.4. *Every compact $S \in \mathcal{L}(H_1, H_2)$ attains its operator norm as a singular value: there are unit vectors u, v with $Su = \|S\|\, v$ and $S^*v = \|S\|\, u$.*
[Lean: `SVD.IsCompactOperator.norm_isSingularValue`]

Proof. Take a maximising sequence x_n of unit vectors with $\|Sx_n\| \to \|S\|$; by compactness extract a subsequence with $Sx_{n_k} \to y$, and set $v := \|S\|^{-1}y$, $u' := S^*v$. Cauchy–Schwarz on $\langle u', x_{n_k}\rangle = \langle v, Sx_{n_k}\rangle \to \|S\|$ gives $\|u'\| = \|S\|$; equality in Cauchy–Schwarz forces $x_{n_k} \to u$, so by continuity $Su = y = \|S\|v$. $\square$

Theorem 6.5 (Schmidt representation)**.** *For every compact $S \in \mathcal{L}(H_1, H_2)$ there exist orthonormal sequences $(u_k) \subseteq H_1$, $(v_k) \subseteq H_2$ such that*

$$Sx \;=\; \sum_k a_k(S)\, \langle u_k, x\rangle\, v_k \qquad \text{for every } x \in H_1.$$

The singular values are exactly the approximation numbers $a_k(S)$.
[Lean: `SVD.IsCompactOperator.SVD`]

Proof. By iterating Theorem 6.4: peel off the top singular pair (σ_0, u_0, v_0) of S, replace S by $S - \sigma_0\langle u_0, \cdot\rangle v_0$ (which is again compact), and recurse. The variational characterisation of each u_k forces the u_k (and dually the

v_k) to be orthonormal; $\sigma_k \to 0$ follows from compactness; and the partial sums converge to S in operator norm because the n-th residual has norm σ_n. Eckart–Young (Theorem 6.6) identifies $\sigma_k = a_k(S)$. □

Theorem 6.6 (Eckart–Young)**.** *For every approximable $S \in \mathcal{L}(H_1, H_2)$ and every $n \in \mathbb{N}_0$ there exists $L \in \mathcal{L}(H_1, H_2)$ with* $\operatorname{rank}(L) \le n$ *and* $\|S - L\| = a_n(S)$*.*

[Lean: `SVD.IsCompactOperator.truncation_residual_eq_approxNumber`]

Proof. Take the truncated SVD $L = \sum_{k<n} a_k(S)\langle u_k, \cdot\rangle v_k$ from Theorem 6.5, which has rank $\le n$. The residual $S - L = \sum_{k \ge n} a_k(S)\langle u_k, \cdot\rangle v_k$ acts on the orthonormal $(u_k)_{k\ge n}$ with weights $a_k(S) \le a_n(S)$, so its operator norm is exactly $a_n(S)$; no rank-$\le n$ operator does better, since $a_n(S)$ is by definition the infimum of $\|S - L'\|$ over such L'. □

Theorem 6.7. *For every compact $S \in \mathcal{L}(H_1, H_2)$ and every $n \in \mathbb{N}_0$ there exist contractions $A \in \mathcal{L}(\ell_2^{n+1}, H_1)$ and $B \in \mathcal{L}(H_2, \ell_2^{n+1})$ such that, on every basis vector e_k of ℓ_2^{n+1},*

$$(B \circ S \circ A)(e_k) \ = \ a_k(S)\, e_k, \qquad 0 \le k \le n.$$

Equivalently, $B \circ S \circ A$ is the diagonal operator $\operatorname{diag}(a_0(S), \dots, a_n(S))$ *on ℓ_2^{n+1}.*

[Lean: `SVD.IsCompactOperator.diagonalFactorisation`]

Proof. Using the top $n+1$ singular pairs of Theorem 6.5, let $A \in \mathcal{L}(\ell_2^{n+1}, H_1)$ send the basis vector e_k to the right singular vector u_k, and let $B \in \mathcal{L}(H_2, \ell_2^{n+1})$ send y to $\sum_{k\le n}\langle v_k, y\rangle e_k$, i.e. project onto the span of the left singular vectors. Both are contractions because (u_k) and (v_k) are orthonormal. Then $Su_k = a_k(S)\, v_k$ gives $(B \circ S \circ A)e_k = a_k(S)\, e_k$ for $0 \le k \le n$. □

This factorisation needs only the top $n + 1$ singular pairs — a finite truncation, a sibling of the full SVD rather than a corollary — but, pinning the *exact* values a_k, it requires S compact.

Theorem 6.8. *For* any *bounded $S \in \mathcal{L}(H_1, H_2)$ between Hilbert spaces and any real c with $0 \le c < a_n(S)$, there are contractions $A \in \mathcal{L}(\ell_2^{n+1}, H_1)$, $B \in \mathcal{L}(H_2, \ell_2^{n+1})$ with $B \circ S \circ A = c \cdot \mathrm{id}_{\ell_2^{n+1}}$.*

[Lean: `SVD.exists_scalar_factorisation`]

Proof. Since $c < a_n(S)$, no rank-$\leq n$ operator approximates S to within c, so there is an $(n+1)$-dimensional subspace $M \subseteq H_1$ on which $\|Sx\| \geq c\,\|x\|$. Let A embed ℓ_2^{n+1} isometrically onto M; then $S \circ A$ is bounded below by c, so $c\,(S \circ A)^{-1}$ is a contraction on its range, and composing it with the orthogonal projection onto that range gives a contraction B with $B \circ S \circ A = c \cdot \mathrm{id}_{\ell_2^{n+1}}$. No compactness is used. $\square$

Unlike the diagonal factorisation this holds for *arbitrary* bounded S (no compactness): the strict inequality $c < a_n(S)$ needs only an $(n+1)$-dimensional subspace on which $\|Sx\| \geq c\|x\|$, which exists for every bounded operator.

6.3 Compact $\Leftrightarrow$ approximable on Hilbert spaces

Theorem 6.9. *Every compact $S \in \mathcal{L}(H_1, H_2)$ between Hilbert spaces is approximable.*
[Lean: `SVD.IsCompactOperator.isApproximable`]

Proof. The Schmidt representation (Theorem 6.5) gives singular values $\sigma_n \to 0$ which, by Eckart–Young (Theorem 6.6), equal the approximation numbers $a_n(S)$; hence $a_n(S) \to 0$. Equivalently, the SVD truncations are finite-rank operators converging to S in operator norm. $\square$

Theorem 6.10. *On Hilbert spaces, S is approximable iff S is compact.*
[Lean: `SVD.isApproximable_iff_isCompactOperator`]

Proof. Combine the two implications: approximable $\Rightarrow$ compact is Theorem 6.3, and compact $\Rightarrow$ approximable is Theorem 6.9. $\square$

6.4 All s-numbers equal the singular values in finite dimension

Mathlib defines the *singular values* of a linear map S between finite-dimensional Hilbert spaces as $\sigma_n(S) := \sqrt{\lambda_n(S^*S)}$, where $\lambda_0 \geq \lambda_1 \geq \ldots$ are the eigenvalues of S^*S in non-increasing order. The following theorem identifies them with *every* s-number sequence at once.

Theorem 6.11. *Let H_1, H_2 be finite-dimensional Hilbert spaces over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. For every s-number sequence s, every $S \in \mathcal{L}(H_1, H_2)$ and every $n \in \mathbb{N}_0$, we have*

$$s_n(S) \;=\; \sigma_n(S),$$

where $\sigma_n(S)$ are Mathlib's eigenvalue-defined singular values.

[Lean: `SNumbers.sn_eq_singularValues_of_finiteDimensional` `SNumbers.project_singularValues_eq`]

Proof. The Hilbert-space coincidence (Theorem 4.1) reduces $s_n(S)$ to $a_n(S)$, so it suffices to identify $a_n(S)$ with $\sigma_n(S)$. The operator S is compact (finite-dimensional domain), so the Schmidt representation (Theorem 6.5) applies and its singular values equal the approximation numbers by Eckart–Young. On the other hand, the eigenvalues of S^*S are exactly the squares of the Schmidt singular values (with the singular vectors as an eigenbasis), so Mathlib's $\sigma_n(S) = \sqrt{\lambda_n(S^*S)}$ coincides with the n-th Schmidt singular value; for $n \geq \dim H_1$ both sides vanish. $\square$

Chapter 7

Inequalities between s-numbers

Beyond the qualitative sandwich $h_n(S) \le s_n(S) \le a_n(S)$ (Theorem 4.2), one can compare the individual example sequences *quantitatively*; see [9] for an overview. This chapter formalises two such comparisons for operators between Banach spaces over $\mathbb{K}$:

- the *forward* bound $a_n(S) \le (1+\sqrt{n}) \min(c_n(S), d_n(S))$ — the largest s-number exceeds the Gelfand and Kolmogorov numbers by at most a factor $1+\sqrt{n}$ (classical; see [6] and [8, §2.10]);
- the *reverse* bound, the *maximal difference theorem* $a_n(S) \le e\,(n+1)\,h_n(S)$ — the maximal possible difference between the largest s-numbers, the approximation numbers, and the smallest ones, the Hilbert numbers, is a factor linear in n [10]. Its consequence for two arbitrary s-number sequences is Corollary 7.14, the conjecture of Carl and Pietsch [1]; specialising to the Gelfand, Kolmogorov and Bernstein numbers gives Corollary 7.15, the conjecture of Mityagin and Henkin [5] — both up to the constant e.

Everything in this chapter is proved in `Inequalities.lean` and `MaxDifference.lean` (the bordered-determinant lemma, a pure matrix fact, in `Determinant.lean`). The two classical projection theorems that drive the forward bound (Garling–Gordon and Kadets–Snobar) are proved, for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ at once, by the John's-ellipsoid development of Section 7.1, whose analytic core is the John decomposition of identity (Theorem 7.2).

7.1 John's ellipsoid and the projection theorems

The sharp constant $\sqrt{n}$ in both classical projection theorems (Theorems 7.3 and 7.4) rests on *John's ellipsoid theorem* [3], which is absent from Mathlib. It is developed in `John.lean` over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ at once, modelling an inscribed ellipsoid as the image $T(B_2)$ of the Euclidean unit ball under a linear map and maximising its volume $|\det T|$. Here a *body seminorm* on $\mathbb{K}^k$ is a continuous seminorm p, thought of as the gauge of a convex body; the hypothesis that it is *equivalent to the Euclidean norm* means $c\,\|x\| \le p(x) \le C\,\|x\|$ for constants $0 < c \le C$. The general-purpose ingredients of the analytic core (closedness of convex hulls of compact sets in finite dimension, the supporting-vector form of Hahn–Banach dominated by a seminorm, trace duality, and a Weierstrass-type product bound) are collected separately, as candidates for Mathlib.

Theorem 7.1. *For a body seminorm p on $\mathbb{K}^k$ equivalent to the Euclidean norm, among the* feasible *operators (those T with $p(Tu) \le \|u\|$ for all u) there is one, T_0, of maximal $|\det|$, and it is invertible ($\det T_0 \neq 0$).*
[Lean: `John.exists_maxVolume`]

Proof. The feasible set is closed (each $T \mapsto p(Tu)$ is continuous) and bounded (the lower norm bound forces $\|T\| \le 1/c$), hence compact; $|\det|$ is continuous, so it attains a maximum. A small multiple of the identity is feasible and has nonzero determinant, so the maximiser does too. □

Theorem 7.2 (John decomposition of identity)**.** *In John position — the identity is a maximal-volume feasible operator for the body seminorm q — the identity is a positive combination of the rank-one projections onto contact points:*

$$\sum_i c_i \langle u_i, x\rangle\, u_i = x \quad (x \in \mathbb{K}^k), \qquad u_i \in \mathrm{Contact}(q),\ c_i \ge 0,\ \sum_i c_i = k,$$

where a contact point is a unit vector u with $\mathrm{Re}\langle x, u\rangle \le q(x)$ for all x.
[Lean: `John.john_decomposition`]

Proof. The classical core of John's theorem, from first-order optimality of the maximal-volume ellipsoid. If $k^{-1}\mathrm{id}$ lay outside the convex hull of the contact rank-ones — a compact convex set of operators, whose compactness uses that convex hulls of compact sets are compact in finite dimension — then geometric Hahn–Banach separation, followed by trace duality, would produce a self-adjoint trace-zero direction H with $\mathrm{Re}\langle u, Hu\rangle < 0$ at every contact point. The perturbation $(1-\rho)^{-1}(\mathrm{id} + tH)$ then stays feasible for small $t > 0$ (the feasibility slack grows linearly in t, using that every touching point is a contact point via a seminorm-dominated Hahn–Banach argument), while its determinant loses only $O(t^2)$ because $\operatorname{tr} H = 0$; so $|\det| > 1$, contradicting maximality. Hence $k^{-1}\mathrm{id}$ lies in the hull, and Carathéodory yields the finite combination. □

Theorem 7.3 (Kadets–Snobar)**.** *Let Y be a normed $\mathbb{K}$-space and $V \subseteq Y$ a subspace of dimension $\leq n$. Then there is a bounded projection $P : Y \to Y$ ($P \circ P = P$) with $\operatorname{range} P = V$ and $\|P\| \leq \sqrt{n}$.*

[Lean: `SNumbers.exists_projection_range_eq_of_rank_le` `John.exists_projection`]

Proof. A classical result [4], see also [8, 1.5.5]. Put V in John position via the maximal-volume ellipsoid (Theorem 7.1): the decomposition of identity (Theorem 7.2) gives contact points u_i, unit vectors $v_i = Mu_i \in V$, and functionals $\varphi_i = \langle u_i, M^{-1}\cdot\rangle$ of norm ≤ 1, which Hahn–Banach extends to $g_i \in Y^*$ without increasing the norm. Then $P = \sum_i c_i\, g_i \otimes v_i$ is the identity on V, hence a projection onto V, and weighted Cauchy–Schwarz with $\sum_i c_i = \dim V$ gives $\|P\| \leq \sqrt{\dim V} \leq \sqrt{n}$. □

Theorem 7.4 (Garling–Gordon)**.** *Let X be a normed $\mathbb{K}$-space and $M \subseteq X$ a closed subspace of codimension $\leq n$. Then for every $\varepsilon > 0$ there is a bounded projection $P : X \to X$ ($P \circ P = P$) with $\ker P = M$ and $\|P\| \leq \sqrt{n} + \varepsilon$.*

[Lean: `SNumbers.exists_projection_ker_eq_of_codim_le`
`John.exists_projection_ker`]

Proof. The dual of Kadets–Snobar (Theorem 7.3) [2], see also [8, 1.7.17]. Run the John development (Theorems 7.1 and 7.2) on the finite-dimensional dual $(X/M)^*$: the contact points are represented by vectors $w_i \in X/M$ with $\|w_i\| \leq 1$ (obtained by solving the single-dual equation $\Phi^{\mathrm{flip}} w_i = \mathrm{toDual}\, u_i$, avoiding the topological double dual), each lifted to $x_i \in X$ with $\|x_i\| < 1 + \varepsilon'$ (the quotient norm is an infimum — the sole source of ε). Then $Py = \sum_i c_i\, (\Phi u_i)(\pi y)\, x_i$ has kernel M and $\|P\| \leq \sqrt{\operatorname{codim} M} + \varepsilon \leq \sqrt{n} + \varepsilon$. The exact bound $\sqrt{n}$ (removing ε) needs a weak-$*$ limit argument; it is not needed here, since the applications recover the sharp constant by letting $\varepsilon \to 0$. □

7.2 Approximation versus Gelfand and Kolmogorov numbers

Theorem 7.5. *For $S \in \mathcal{L}(X, Y)$ between Banach spaces, $a_n(S) \le (1 + \sqrt{n})\, c_n(S)$.*

[Lean: `SNumbers.approximationNumber_le_sqrt_mul_gelfandNumber`]

Proof. Fix a closed $M \subseteq X$ of codimension $\le n$ with $\|S|_M\|$ near $c_n(S)$. By Garling–Gordon (Theorem 7.4), for each $\varepsilon > 0$ choose a projection P with $\ker P = M$ and $\|P\| \le \sqrt{n} + \varepsilon$. Then $L := S \circ P$ has rank $\le \operatorname{codim} M \le n$, and $S - L = S \circ (\mathrm{id} - P)$ has range inside M (since $\operatorname{range}(\mathrm{id} - P) = \ker P = M$), so $\|S - L\| \le \|S|_M\| \, \|\mathrm{id} - P\| \le (1 + \sqrt{n} + \varepsilon)\, \|S|_M\|$. Letting $\varepsilon \to 0$ gives $a_n(S) \le (1 + \sqrt{n})\, \|S|_M\|$; taking the infimum over M gives the claim. □

Theorem 7.6. *For $S \in \mathcal{L}(X, Y)$ between Banach spaces, $a_n(S) \le (1 + \sqrt{n})\, d_n(S)$.*

[Lean: `SNumbers.approximationNumber_le_sqrt_mul_kolmogorovNumber`]

Proof. Dual to the Gelfand half (Theorem 7.5). Fix $V \subseteq Y$ of dimension $\le n$ with $\|\pi_V \circ S\|$ near $d_n(S)$; by Kadets–Snobar (Theorem 7.3) choose a projection P onto V with $\|P\| \le \sqrt{n}$. Then $L := P \circ S$ has rank $\le n$, and $S - L = (\mathrm{id} - P) \circ S$ factors through the quotient Y/V (because $\mathrm{id} - P$ vanishes on V), so $\|S - L\| \le \|\mathrm{id} - P\| \, \|\pi_V \circ S\| \le (1 + \sqrt{n})\, d_n(S)$ after taking the infimum. □

Theorem 7.7. *For $S \in \mathcal{L}(X, Y)$ between Banach spaces,*

$$a_n(S) \;\le\; (1 + \sqrt{n})\, \min\bigl(c_n(S), d_n(S)\bigr).$$

[Lean: `SNumbers.approximationNumber_le_sqrt_mul_min`]

Proof. Immediate from the two halves (Theorems 7.5 and 7.6) by case analysis on which of $c_n(S)$, $d_n(S)$ is the smaller. □

7.3 The maximal difference theorem

The maximal difference theorem $a_n \le e\,(n+1)\, h_n$ is due to Ullrich [10]. It is driven by the *determinant quantities* of Definition 7.8, which are squeezed between the two quantities of interest: they do not decay faster than the approximation numbers allow (Lemma 7.11, via the explicit rank-k approximant $L = SA(BSA)^{-1}BS$), and consecutive ratios are bounded by the Hilbert numbers (Lemma 7.12, via contractive compressions).

Definition 7.8 (Determinant quantities)**.** Call a pair (A, B) *admissible* for $S \in \mathcal{L}(X, Y)$ and $k \in \mathbb{N}_0$ if $A : \ell_2^k \to X$ and $B : Y \to \ell_2^k$ are contractions. For such a pair $B \circ S \circ A$ is an endomorphism of the k-dimensional Hilbert space ℓ_2^k, so its determinant is a scalar, and

$$\Delta_k(S) \ := \ \sup\big\{ \,|\det(B \circ S \circ A)| : (A, B) \text{ admissible for } S \text{ and } k \,\big\}.$$

[Lean: `SNumbers.detNumber`]

One has $\Delta_0(S) = 1$ (the determinant on the trivial space is 1) and $\Delta_k(S) \le \|S\|^k$, via $\prod_i a_i(T) = |\det T|$ (Lemma 7.9) and $a_i(T) \le \|T\| \le \|S\|$.

Lemma 7.9. *For an endomorphism T of a finite-dimensional inner-product space, $\prod_{k<\dim} a_k(T) = |\det T|$.*
[Lean: `SNumbers.prod_approximationNumber_eq_norm_det`]

Proof. The approximation numbers coincide with the singular values $\sigma_k(T)$ (Theorem 6.11), and $|\det T| = \prod_k \sigma_k(T)$ is linear algebra: $|\det T|^2 = \det(T^*T) = \prod_k \lambda_k(T^*T) = \prod_k \sigma_k(T)^2$. □

The bordered-determinant identity is the elementary column-operation form of the Schur determinant formula; it needs no invertibility of the top-left block.

Lemma 7.10 (Bordered determinant)**.** *Let R be a commutative ring, M a $(k+1) \times (k+1)$ matrix over R and $w \in R^k$ such that, above the corner, the last column of M is the combination of the first k columns with coefficients w, i.e. $M_{i,\mathrm{last}} = \sum_j M_{ij} w_j$ for $i < k$. Then*

$$\det M \ = \ \Big(M_{\mathrm{last,last}} - \sum_j M_{\mathrm{last},j}\, w_j\Big) \cdot \det M',$$

where M' is the top-left $k \times k$ block of M.
[Lean: `Matrix.det_eq_corner_mul_det_submatrix`]

Proof. Subtracting from the last column the combination of the other columns with coefficients w leaves the determinant unchanged and clears the column above the corner; Laplace expansion along it gives the claim. □

Lemma 7.11. *For every $S \in \mathcal{L}(X, Y)$ and $k \in \mathbb{N}_0$,*

$$\frac{k^k}{(k+1)^{k+1}}\, a_k(S)\, \Delta_k(S) \ \le \ \Delta_{k+1}(S).$$

[Lean: `SNumbers.approximationNumber_mul_detNumber_le_detNumber_succ`]

Proof. Fix an admissible pair (A, B) for $\Delta_k(S)$ and put $T := BSA$; if $\det T = 0$ there is nothing to prove. Otherwise T is invertible, so $L := SAT^{-1}BS$ factors through ℓ_2^k, hence has rank at most k, hence $\|S - L\| \geq a_k(S)$ by the definition of the approximation numbers. Pick $x \in B_X$ with $\|(S - L)x\|$ almost $\|S - L\|$ and a Hahn–Banach functional $b \in B_{Y^*}$ norming $(S - L)x$, and border the pair by them: $A_0(\xi \oplus \tau) := \lambda A\xi + \mu \tau x$ and $B_0 y := (\lambda By) \oplus (\mu b(y))$ with $\lambda^2 + \mu^2 = 1$. Contractivity of A_0 is the two-term Cauchy–Schwarz inequality $\lambda a + \mu b \leq \sqrt{a^2 + b^2}$, and $\|B_0 y\|^2 = \lambda^2\|By\|^2 + \mu^2|b(y)|^2 \leq \|y\|^2$. Above the corner, the last column of the matrix of B_0SA_0 is the combination of the first k columns with coefficients $(\mu/\lambda)\,\zeta$, where $\zeta := T^{-1}B(Sx)$, so Lemma 7.10 gives

$$\det(B_0SA_0) \;=\; \lambda^{2k}\mu^2\, b\big((S - L)x\big)\det T.$$

The weights $\lambda^2 = k/(k + 1)$, $\mu^2 = 1/(k + 1)$ maximise $\lambda^{2k}\mu^2$, whose value is then exactly $k^k/(k + 1)^{k+1}$; taking the supremum over (A, B) finishes. Note that no near-maximiser has to be selected — *every* admissible pair is improved. □

In particular, if $a_n(S) > 0$ then $\Delta_k(S) > 0$ for all $k \leq n+1$, by induction from $\Delta_0(S) = 1$ and $a_k(S) \geq a_n(S)$.

Lemma 7.12. $\Delta_{n+1}(S) \leq h_n(S)\,\Delta_n(S)$.
[Lean: `SNumbers.detNumber_succ_le_hilbertNumber_mul`]

Proof. For admissible (A, B) and $T := BSA$ on ℓ_2^{n+1}, $|\det T| = \prod_{k\leq n} a_k(T)$ (Lemma 7.9). The last factor satisfies $a_n(T) \leq \|B\|\|A\|\,h_n(S) \leq h_n(S)$. For the top n factors, the diagonal factorisation $B_1TA_1 = \mathrm{diag}(a_0(T), \ldots, a_n(T))$ through contractions (Theorem 6.7), compressed to the first n coordinates by the coordinate embedding/projection, is itself admissible for $\Delta_n(S)$, whence $\prod_{k<n} a_k(T) \leq \Delta_n(S)$. Again every admissible pair is bounded, so the bound passes to the supremum $\Delta_{n+1}(S)$. □

Theorem 7.13 (Maximal difference theorem)**.** *For every $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,*

$$a_n(S) \;\leq\; \frac{(n + 1)^{n+1}}{n^n}\, h_n(S) \;\leq\; e\,(n + 1)\, h_n(S).$$

[Lean: `SNumbers.approximationNumber_le_mul_hilbertNumber`
`SNumbers.approximationNumber_le_e_mul_hilbertNumber`]

Proof. If $a_n(S) = 0$ there is nothing to prove. Otherwise chain the growth lemma (Lemma 7.11) at $k = n$ with the upper bound (Lemma 7.12): $\frac{n^n}{(n+1)^{n+1}} a_n(S)\,\Delta_n(S) \le \Delta_{n+1}(S) \le h_n(S)\,\Delta_n(S)$, and cancel $\Delta_n(S) > 0$. Finally $(n+1)^{n+1}/n^n = (n+1)(1+1/n)^n \le e\,(n+1)$ from $1 + x \le e^x$ at $x = 1/n$. ☐

Since the Hilbert numbers are the smallest and the approximation numbers the largest s-number sequence, the theorem bounds *any* s-number sequence by any other: the maximal possible difference between two s-number sequences is a factor linear in n. This is the conjecture of Carl and Pietsch [1], proved up to the constant e in [10].

Corollary 7.14. *For all s-number sequences s, t, every $S \in \mathcal{L}(X, Y)$ and every $n \in \mathbb{N}_0$,*

$$s_n(S) \;\le\; e\,(n+1)\,t_n(S).$$

[Lean: `SNumbers.sn_le_e_mul_tn`]

Proof. By the sandwich theorem (Theorem 4.2) the Hilbert numbers are the smallest and the approximation numbers the largest s-number sequence, so Theorem 7.13 gives $s_n(S) \le a_n(S) \le e\,(n+1)\,h_n(S) \le e\,(n+1)\,t_n(S)$. ☐

Taking $t = b$ and $s = c$ resp. $s = d$ gives the *Mityagin–Henkin conjecture* [5], again up to the constant e.

Corollary 7.15 (Mityagin–Henkin)**.** *For every $S \in \mathcal{L}(X, Y)$ and every $n \in \mathbb{N}_0$,*

$$\max\bigl(c_n(S), d_n(S)\bigr) \;\le\; e\,(n+1)\,b_n(S).$$

[Lean: `SNumbers.max_gelfandNumber_kolmogorovNumber_le_e_mul_bernsteinNumber`]

Proof. Apply Corollary 7.14 with $t := b$ and $s := c$ resp. $s := d$ — both are s-number sequences — and take the maximum. ☐

The factor $n+1$ is order-optimal: the inclusion $I\colon \ell_1 \to \ell_\infty$ of Section 8.3 satisfies $c_n(I) \ge \frac{1}{2}$ while $h_n(I) = (n+1)^{-1}$, so no bound $a_n \le \varphi(n)\,h_n$ can hold with $\varphi(n) = o(n)$ (Corollary 8.15).

Chapter 8

Examples

This chapter computes the s-numbers of three concrete families of operators. The finite-dimensional examples — the identity between ℓ_p-spaces and the diagonal operators — are the standard test cases of the theory; the inclusion $\ell_1 \to \ell_\infty$ shows that the linear factor $n+1$ in the maximal difference theorem (Theorem 7.13) cannot be improved in order. The chapter is a leaf of the development: it consumes the general theory and nothing depends on it. Throughout, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, and ℓ_p^m denotes $\mathbb{K}^m$ with the ℓ_p-norm.

8.1 The identity $\mathrm{id} : \ell_q^m \to \ell_p^m$

Theorem 8.1. *For $1 \le p \le q < \infty$ and $m \ge 1$,*

$$\big\|\mathrm{id} : \ell_q^m \to \ell_p^m\big\| \;=\; m^{\,1/p-1/q}.$$

[Lean: `SNumbers.norm_idEmbed`]

Proof. The upper bound is the power-mean (Hölder) inequality $\|x\|_p \le m^{1/p-1/q}\|x\|_q$; the lower bound tests the all-ones vector, whose ℓ_q-norm is $m^{1/q}$ and whose ℓ_p-norm is $m^{1/p}$. □

Theorem 8.2. *For $1 \le p \le q < \infty$ and $n < m$,*

$$a_n\big(\mathrm{id} : \ell_q^m \to \ell_p^m\big) \;=\; (m-n)^{\,1/p-1/q}.$$

[Lean: `SNumbers.approximationNumber_idEmbed_eq`]

Proof. For the upper bound, truncate to the first n coordinates: the residual is the identity on the remaining $m-n$ coordinates, of norm $(m-n)^{1/p-1/q}$ by (the proof of) Theorem 8.1. For the lower bound, a rank-$\le n$ operator L has kernel of dimension at least $m-n$, and a flatness (pigeonhole) argument produces a nonzero x in the kernel with $\|x\|_p \ge (m-n)^{1/p-1/q}\|x\|_q$; on such an x the error $\mathrm{id}-L$ acts as the identity, so $\|\mathrm{id}-L\| \ge (m-n)^{1/p-1/q}$. ☐

By the sandwich theorem (Theorem 4.2), every s-number sequence obeys $s_n(\mathrm{id}) \le (m-n)^{1/p-1/q}$.

Proposition 8.3. *For every strict s-number sequence s, every $1 \le p \le \infty$ and $n < m$, we have $s_n(\mathrm{id}_{\ell_p^m}) = 1$.*

[Lean: `SNumbers.sn_id_piLp`]

Proof. This is exactly the strict normalisation (S5′) applied to $X = \ell_p^m$, whose dimension is $m > n$. ☐

8.2 Diagonal operators

For $\sigma = (\sigma_0, \dots, \sigma_{m-1}) \in \mathbb{K}^m$ let D_σ denote the diagonal operator

$$D_\sigma : (x_i)_i \longmapsto (\sigma_i\, x_i)_i.$$

Throughout this section, σ is *non-increasing in modulus*, $|\sigma_0| \ge |\sigma_1| \ge \cdots \ge |\sigma_{m-1}|$, and $n < m$.

Theorem 8.4 (Same exponent)**.** *Let $1 \le p \le \infty$ and $D_\sigma : \ell_p^m \to \ell_p^m$. Then $\|D_\sigma\| = \max_i |\sigma_i|$, and for every* strict *s-number sequence s (in particular for a_n, b_n, c_n, d_n),*

$$s_n(D_\sigma) \;=\; |\sigma_n|.$$

[Lean: `SNumbers.norm_DiagCLM` `SNumbers.sn_DiagCLM_eq`]

Proof. The norm is attained at the coordinate carrying $\max_i |\sigma_i| = |\sigma_0|$. Upper bound: truncating σ to its first n entries gives a rank-$\le n$ approximant whose residual is the diagonal operator with entries $(\sigma_k)_{k\ge n}$, of norm $|\sigma_n|$; hence $s_n(D_\sigma) \le a_n(D_\sigma) \le |\sigma_n|$ for *every* s-number sequence. Lower bound for strict s: since $|\sigma_k| \ge |\sigma_n|$ for $k \le n$, the identity of ℓ_p^{n+1} factors as $\mathrm{id} = B \circ D_\sigma \circ A$ with the coordinate embedding A (a contraction) and B the coordinate projection followed by the diagonal operator with entries σ_k^{-1}, of norm $\le |\sigma_n|^{-1}$; then (S3) and the strict normalisation (Proposition 8.3) give $1 = s_n(\mathrm{id}) \le |\sigma_n|^{-1} s_n(D_\sigma)$. ☐

The Hilbert numbers are not strict, and satisfy only $h_n(D_\sigma) \le |\sigma_n|$; equality can fail for $p \neq 2$.

Theorem 8.5 (Mixed exponents $p < q$)**.** *Let $1 \le p < q < \infty$ and define r by $\frac{1}{r} = \frac{1}{p} - \frac{1}{q}$. For $D_\sigma : \ell_q^m \to \ell_p^m$,*

$$a_n(D_\sigma) \;=\; \Big(\sum_{k=n}^{m-1} |\sigma_k|^r\Big)^{1/r}, \qquad \textit{in particular} \qquad \|D_\sigma\| = \|\sigma\|_{\ell_r}.$$

[Lean: `SNumbers.norm_DiagCLMpq_eq` `SNumbers.approximationNumber_DiagCLMpq_eq`]

Proof. Upper bound: truncate to the first n coordinates; the residual diagonal has norm equal to the ℓ_r-norm of its entries, by Hölder's inequality applied with the exponent pair $(q/r', \dots)$ and attained at the explicitly weighted vector $x_k = |\sigma_k|^{r/q}$. Lower bound: for a rank-$\le n$ operator L, a weighted flatness argument on $\ker L$ (the weighted analogue of the pigeonhole step in Theorem 8.2) produces a kernel vector on which $\mathrm{id} - L$ retains the full tail ℓ_r-norm. The case $n = 0$, $L = 0$ yields the operator norm. □

By the sandwich theorem (Theorem 4.2) the upper bound $s_n(D_\sigma) \le \big(\sum_{k \ge n} |\sigma_k|^r\big)^{1/r}$ holds for every s-number sequence, and it needs no ordering of the diagonal.

Theorem 8.6 (Reverse regime $q \le p$)**.** *Let $1 \le q \le p \le \infty$ (including $p = \infty$). For $D_\sigma : \ell_q^m \to \ell_p^m$,*

$$\|D_\sigma\| = \max_i |\sigma_i|, \qquad s_n(D_\sigma) \le |\sigma_n| \quad \textit{for every s-number sequence } s.$$

[Lean: `SNumbers.norm_DiagCLMpq_iSup` `SNumbers.sn_DiagCLMpq_le_of_exponent_ge`]

Proof. Since $q \le p$, the identity $\ell_q^m \to \ell_p^m$ is a contraction, and D_σ factors as this contraction after the same-exponent diagonal operator on ℓ_q^m; the ideal property (S3) transfers both bounds from Theorem 8.4. The lower bound for the norm again tests a single coordinate. □

For $q < p$ the upper bound $s_n(D_\sigma) \le |\sigma_n|$ need not be sharp: the exact values have no elementary closed form in this regime.

8.3 The inclusion $\ell_1 \to \ell_\infty$

This section computes both sides of the maximal difference theorem for the natural inclusion

$$I\colon \ell_1 \to \ell_\infty, \qquad (x_j)_j \mapsto (x_j)_j,$$

a contraction which is *not* compact. Everything here is formalised in `IdentityL1Linfty.lean`, with the two little Grothendieck bounds in `LittleGrothendieck.lean` and the product estimate, a general Hilbert-space result, in `SVD.lean`.

8.3.1 Ingredients

Two inputs are needed for the Hilbert numbers of I. The first pair of lemmas bounds a Hilbert–Schmidt-type sum by an operator norm; the second is an estimate on the leading singular values of a product.

Lemma 8.7 (Sign averaging)**.** *Let w_j ($j \in J$, J finite) be vectors in an inner product space and $M \in \mathbb{R}$. If $\|\sum_{j\in J} \varepsilon_j w_j\| \le M$ for every choice of signs $\varepsilon \in \{\pm 1\}^J$, then $\sum_{j\in J} \|w_j\|^2 \le M^2$.*

[Lean: `SNumbers.sum_norm_sq_le_sq_of_signedSum_le`]

Proof. Summing over all $2^{|J|}$ sign patterns, the cross terms cancel:

$$\sum_{\varepsilon\in\{\pm1\}^J} \Big\|\sum_{j\in J} \varepsilon_j w_j\Big\|^2 = 2^{|J|} \sum_{j\in J} \|w_j\|^2,$$

proved by induction on J, pairing the two extensions of each sign pattern to a new index by the parallelogram law. Each summand on the left is at most M^2, and there are $2^{|J|}$ of them. □

Lemma 8.8 (Little Grothendieck bounds)**.** *Let H be a Hilbert space and J a finite set of indices.*

1. *For every $B \in \mathcal{L}(\ell_\infty, H)$, $\displaystyle\sum_{j\in J} \|Be_j\|^2 \le \|B\|^2$.*
2. *For every $A \in \mathcal{L}(H, \ell_1)$ with rows $w_j \in H$, i.e. $\langle w_j, x\rangle = (Ax)_j$ for all x, $\displaystyle\sum_{j\in J} \|w_j\|^2 \le \|A\|^2$.*

In Hilbert–Schmidt language: both operators are Hilbert–Schmidt with $\|\cdot\|_{HS} \le \|\cdot\|$.

[Lean: `SNumbers.sum_norm_sq_apply_single_le` `SNumbers.sum_norm_sq_row_le`]

Proof. (1) The vectors $\sum_{j\in J} \varepsilon_j e_j$ all have ℓ_∞-norm 1, so $\|\sum_j \varepsilon_j B e_j\| \le \|B\|$ and Lemma 8.7 applies with $M = \|B\|$.

(2) For a signed sum $z := \sum_{j\in J} \varepsilon_j w_j$ the rows reproduce the signs on the coordinates of Az:

$$\|z\|^2 = \mathrm{Re}\langle z, z\rangle = \mathrm{Re} \sum_{j\in J} \varepsilon_j (Az)_j \le \sum_{j\in J} |(Az)_j| \le \|Az\|_1 \le \|A\|\,\|z\|,$$

so $\|z\| \le \|A\|$, and Lemma 8.7 applies with $M = \|A\|$. This is the little Grothendieck bound for the dual operator $\ell_\infty = \ell_1' \to H$, formulated through the rows so that no adjoint has to be constructed. ☐

The second ingredient is a consequence of the Schmidt representation (Theorem 6.5): the quantitative form of the classical statement that the nuclear norm of a product is at most the product of the Hilbert–Schmidt norms of its factors [8, 2.11.23]. The two Hilbert–Schmidt bounds enter as hypotheses, so neither the Hilbert–Schmidt nor the nuclear norm has to be defined.

Lemma 8.9. *Let $S \in \mathcal{L}(H_1, H_2)$ be described by* rows $w_j \in H_1$, *in the sense that $\|Sx\|^2 = \sum_j |\langle w_j, x\rangle|^2$ for all x, and suppose $\sum_{j\in J} \|w_j\|^2 \le c$ for every finite J. Then $\sum_{k\in s} \|Su_k\|^2 \le c$ for every finite s and every family (u_k) of pairwise orthogonal vectors of norm 1 or 0.*

[Lean: `SVD.sum_norm_sq_apply_le_of_rows`]

Proof. Interchange the two summations and apply Bessel's inequality $\sum_k |\langle u_k, w_j\rangle|^2 \le \|w_j\|^2$ for each fixed j, then sum over j and use the row bound. ☐

Theorem 8.10. *Let $T = T_2 T_1$ be compact, with $T_1 \in \mathcal{L}(H_1, H_2)$ and $T_2 \in \mathcal{L}(H_2, H_3)$, and let $a, b \ge 0$ be such that*

$$\sum_{k\in s} \|T_1 u_k\|^2 \le a^2 \qquad \textit{and} \qquad \sum_{k\in s} \|T_2^* v_k\|^2 \le b^2$$

for all finite s and all orthonormal(-or-zero) families (u_k) in H_1 and (v_k) in H_3. Then

$$(n+1)\, a_n(T) \;\le\; \sum_{k=0}^{n} \sigma_k(T) \;\le\; b\, a \qquad (n \in \mathbb{N}_0).$$

[Lean: `SVD.mul_approximationNumber_le_of_factorization`]

Proof. Take a Schmidt representation of T with singular values σ_k and singular vectors u_k, v_k; by Eckart–Young $\sigma_n = a_n(T)$, and since σ is antitone, $(n+1)\sigma_n \le \sum_{k\le n} \sigma_k$. For each k with $\sigma_k \ne 0$ the vector v_k is a unit vector and $Tu_k = \sigma_k v_k$, whence

$$\sigma_k = \mathrm{Re}\langle v_k, Tu_k\rangle = \mathrm{Re}\langle T_2^* v_k, T_1 u_k\rangle \le \|T_2^* v_k\| \, \|T_1 u_k\|.$$

Summing over $k \le n$ and applying Cauchy–Schwarz together with the two hypotheses gives $\sum_{k\le n} \sigma_k \le b\,a$. □

8.3.2 The Gelfand numbers of I

Proposition 8.11. *For all n, $\frac{1}{2} \le c_n(I) \le 1$.*
[Lean: `SNumbers.L1Linf.gelfand_le_one` `SNumbers.L1Linf.half_le_gelfandNumber`]

Proof. The upper bound is $c_n(I) \le \|I\| \le 1$. For the lower bound let $N \subseteq \ell_1$ be closed of codimension $\le n$. Then ℓ_1/N is finite dimensional, so among the classes $[e_i]$ of the unit vectors two are arbitrarily close (Bolzano–Weierstrass). Lifting the difference gives $x \in N$ close to $e_i - e_{i'}$, with $\|x\|_1 \approx 2$ but $\|Ix\|_\infty \approx 1$; hence $\|I|_N\| \ge \frac{1}{2} - \varepsilon$ for every $\varepsilon > 0$. □

8.3.3 The Hilbert numbers of I

Proposition 8.12. $h_n(I) \ge (n+1)^{-1}$ *for all n.*
[Lean: `SNumbers.L1Linf.one_div_le_hilbertNumber`]

Proof. The identity of ℓ_2^{n+1} factors as $\mathrm{id}_{\ell_2^{n+1}} = \mathrm{proj}_\infty \circ I \circ \mathrm{emb}_1$ with $\|\mathrm{emb}_1\|, \|\mathrm{proj}_\infty\| \le \sqrt{n+1}$. By the ideal property of the Hilbert numbers and their normalisation $h_n(\mathrm{id}_{\ell_2^{n+1}}) \ge 1$,

$$1 \le \|\mathrm{proj}_\infty\| \, h_n(I) \, \|\mathrm{emb}_1\| \le (n+1)\, h_n(I).$$

□

Theorem 8.13. $h_n(I) \le (n+1)^{-1}$ *for all n.*
[Lean: `SNumbers.L1Linf.hilbertNumber_le_one_div`]

Proof. Let $A \in \mathcal{L}(\ell_2, \ell_1)$ and $B \in \mathcal{L}(\ell_\infty, \ell_2)$ be non-zero; we must bound $a_n(BIA) \le \|A\| \, \|B\|/(n+1)$. Factorising $I = J_2 J_1$ with the norm-one inclusions $J_1 \colon \ell_1 \to \ell_2$ and $J_2 \colon \ell_2 \to \ell_\infty$, we obtain $T := BIA = T_2 T_1$ with

$T_1 := J_1 A$ and $T_2 := BJ_2$ on ℓ_2. Applying Lemma 8.8 to B and to the rows of A,

$$\|T_2\|_{HS}^2 = \sum_j \|Be_j\|^2 \le \|B\|^2 \quad \text{and} \quad \|T_1\|_{HS}^2 = \sum_j \|\mathrm{row}_j(A)\|^2 \le \|A\|^2,$$

which by Lemma 8.9 gives the two hypotheses of Theorem 8.10 along the singular vectors of T. That theorem yields $(n+1)\, a_n(T) \le \|A\|\,\|B\|$ *provided T is compact.*

For compactness, replace A by its coordinate truncation A_m, of rank $\le m$ and with $\|A_m\| \le \|A\|$; then BIA_m has finite rank, hence is compact (Theorem 6.3), and the estimate above applies to it. Since $\sum_j \|\mathrm{row}_j(A)\|^2 < \infty$ we have $\|\mathrm{row}_j(A)\| \to 0$, so $\|I(A - A_m)\| \to 0$, and subadditivity of the approximation numbers,

$$a_n(BIA) \le a_n(BIA_m) + \|B\|\,\|I(A - A_m)\|,$$

lets $m \to \infty$. Dividing by $\|A\|\,\|B\|$ and taking the supremum over all admissible A, B gives $h_n(I) \le (n+1)^{-1}$. □

Corollary 8.14. *$h_n(I) = (n+1)^{-1}$ for all n.*
[Lean: `SNumbers.L1Linf.hilbertNumber_eq_one_div`]

Proof. Combine Proposition 8.12 and Theorem 8.13. □

8.3.4 Order-optimality of the factor $n + 1$

Corollary 8.15. *For all n,*

$$\frac{n+1}{2}\, h_n(I) \;\le\; c_n(I) \;\le\; a_n(I),$$

whereas the maximal difference theorem gives $a_n \le e\,(n+1)\, h_n$ in general. Hence no bound $a_n \le \varphi(n)\, h_n$ can hold with $\varphi(n) = o(n)$.
[Lean: `SNumbers.L1Linf.mul_hilbertNumber_le_gelfandNumber`]

Proof. By Corollary 8.14, $\frac{n+1}{2}\, h_n(I) = \frac{1}{2}$, which is at most $c_n(I)$ by Proposition 8.11. □

Chapter 9

Injective and surjective s-numbers

Two further structural properties of an s-number sequence, from Pietsch's theory, single out the Gelfand and Kolmogorov numbers. They concern the behaviour of s_n under composition with two distinguished classes of maps. Everything in this chapter is fully formalised in `Injectivity.lean`.

Definition 9.1 (Metric injection)**.** A *metric injection* is a continuous linear map $J : Y \to Z$ that preserves norms: $\|Jy\| = \|y\|$ for all y (an isometric embedding). In particular $\|J\| \leq 1$.
[Lean: `SNumbers.IsMetricInjection`]

Definition 9.2 (Metric surjection)**.** A *metric surjection* is a continuous linear map $Q : W \to X$ that is a contraction ($\|Q\| \leq 1$) and maps the open unit ball *onto* the open unit ball: every x has preimages of norm arbitrarily close to $\|x\|$. This is the quotient-map condition $\|x\| = \inf\{\|w\| : Qw = x\}$.
[Lean: `SNumbers.IsMetricSurjection`]

Lemma 9.3. *Post-composition with a metric injection and pre-composition with a metric surjection both leave the operator norm unchanged:* $\|J \circ T\| = \|T\|$ *and* $\|T \circ Q\| = \|T\|$.
[Lean: `SNumbers.IsMetricInjection.norm_comp`
`SNumbers.IsMetricSurjection.norm_comp`]

Proof. For the injection, $\|(J \circ T)x\| = \|J(Tx)\| = \|Tx\|$ pointwise, so the two operator norms agree. For the surjection, $\|T \circ Q\| \leq \|T\|$ since $\|Q\| \leq 1$; conversely, lifting each x to a preimage w with $\|w\| \approx \|x\|$ gives $\|Tx\| =$

$\|(T \circ Q)w\| \leq \|T \circ Q\| \, \|w\|$, and letting the slack tend to 0 yields $\|T\| \leq \|T \circ Q\|$. $\square$

Definition 9.4 (Injective / surjective s-number sequence)**.** An s-number sequence s is *injective* if post-composition with any metric injection leaves it unchanged, $s_n(J \circ S) = s_n(S)$; it is *surjective* if pre-composition with any metric surjection leaves it unchanged, $s_n(S \circ Q) = s_n(S)$.

[Lean: `SNumbers.Injective SNumbers.Surjective`]

Theorem 9.5. *The Gelfand numbers c_n form an injective s-number sequence: $c_n(J \circ S) = c_n(S)$ for every metric injection J.*

[Lean: `SNumbers.injective_gelfandNumber`]

Proof. The infimum defining c_n ranges over restriction norms $\|S|_M\|$, and $\|(J \circ S)|_M\| = \|J \circ (S|_M)\| = \|S|_M\|$ by Lemma 9.3. The two infimum sets therefore coincide term by term, so the Gelfand numbers agree. $\square$

Theorem 9.6. *The Kolmogorov numbers d_n form a surjective s-number sequence: $d_n(S \circ Q) = d_n(S)$ for every metric surjection Q.*

[Lean: `SNumbers.surjective_kolmogorovNumber`]

Proof. The infimum defining d_n ranges over quotient norms $\|\pi_V \circ S\|$, and $\|\pi_V \circ (S \circ Q)\| = \|(\pi_V \circ S) \circ Q\| = \|\pi_V \circ S\|$ by Lemma 9.3. The two infimum sets coincide term by term, so the Kolmogorov numbers agree. $\square$

Chapter 10

Entropy numbers

The entropy numbers quantify the compactness of an operator directly: they measure how many balls of a given radius are needed to cover the image of the unit ball. They are *not* s-numbers — the rank axiom (S4) and the normalisation (S5) both fail — but they satisfy the remaining properties, in the strong additive and multiplicative form of Theorems 10.3 and 10.4; such sequences are called *pseudo*-s-*number* sequences [7, 12.1.1]. Everything in this chapter is formalised in `Entropy.lean`, over a non-trivially normed field $\mathbb{K}$ unless stated otherwise.

Definition 10.1 (Entropy number)**.** For $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,

$$e_n(S) := \inf\big\{\varepsilon > 0 : S(B_X) \text{ is covered by at most } 2^n \text{ closed } \varepsilon\text{-balls in } Y\big\},$$

where B_X is the closed unit ball of X and the centres of the covering balls are arbitrary points of Y. (The indexing is 0-based, so e_0 allows a single ball.)

[Lean: `SNumbers.entropyNumber`]

Proposition 10.2. *For every $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}_0$,*

$$0 \;\leq\; e_{n+1}(S) \;\leq\; e_n(S) \;\leq\; \|S\|.$$

[Lean: `SNumbers.entropyNumber_antitone SNumbers.entropyNumber_le_norm`]

Proof. Nonnegativity and monotonicity are immediate from the definition (any cover by 2^n balls is a cover by 2^{n+1} balls). For the norm bound, a single ball of radius $\|S\| + \delta$ centred at the origin covers $S(B_X)$, for every $\delta > 0$. $\square$

Theorem 10.3. *For all $S, T \in \mathcal{L}(X,Y)$ and $n, m \in \mathbb{N}_0$,*

$$e_{n+m}(S+T) \;\le\; e_n(S) + e_m(T).$$

[Lean: `SNumbers.entropyNumber_additive` `SNumbers.entropyNumber_add_le`]

Proof. If $S(B_X)$ is covered by 2^n balls of radius ε with centres y_i and $T(B_X)$ by 2^m balls of radius δ with centres z_j, then the 2^{n+m} balls of radius $\varepsilon+\delta$ centred at the sums $y_i + z_j$ cover $(S+T)(B_X)$. □

Taking $m = 0$ and using $e_0(T) \le \|T\|$ (Proposition 10.2) recovers the s-number axiom (S2): $e_n(S+T) \le e_n(S) + \|T\|$.

Theorem 10.4. *Let the norm values of $\mathbb{K}$ be dense in $[0,\infty)$ (as for $\mathbb{R}$ and $\mathbb{C}$). Then for all $S \in \mathcal{L}(X,Y)$, $B \in \mathcal{L}(Y,Z)$ and $n, m \in \mathbb{N}_0$,*

$$e_{n+m}(B \circ S) \;\le\; e_n(B)\, e_m(S).$$

[Lean: `SNumbers.entropyNumber_multiplicative`
`SNumbers.entropyNumber_comp_comp_le`]

Proof. Cover $S(B_X)$ by 2^m balls of radius δ with centres y_i. Each point of $S(B_X)$ in the ball around y_i differs from y_i by a vector u with $\|u\| \le \delta$, which is rescaled as $u = c\,w$ with $\|w\| \le 1$ and $|c|$ close to δ — this is where density of the norm values enters. Covering $B(B_Y)$ by 2^n balls of radius ε then covers each $B(u)$ by 2^n balls of radius $|c|\,\varepsilon$, and combining the two covers yields 2^{n+m} balls of radius close to $\varepsilon\delta$ around $B(S(B_X))$. □

Taking one index to be 0 and using $e_0 \le \|\cdot\|$ twice (Proposition 10.2) gives the s-number axiom (S3): $e_n(B \circ S \circ A) \le \|B\|\, e_n(S)\, \|A\|$.

Remark 10.5 (Entropy numbers are not s-numbers)**.** *Two of Pietsch's axioms fail, and one is not formalised:*

- *(S4) fails: if $e_n(S) = 0$ then $S(B_X)$ is covered by 2^n balls of arbitrarily small radius, hence has at most 2^n points — impossible for $S \ne 0$. So $e_n(S) > 0$ even for finite-rank S.*
- *(S5) fails for $n \ge 2$: covering the Euclidean unit ball of $\mathbb{R}^{n+1}$ by balls of radius < 1 takes only $n+2$ of them, and $2^n \ge n+2$ from $n = 2$ on, so $e_n(\mathrm{id}_{\ell_2^{n+1}}) < 1$.*

- *(S1a), the equality* $e_0(S) = \|S\|$*, holds over* $\mathbb{R}$ *and* $\mathbb{C}$ *by symmetry of* $S(B_X)$*, but fails over general normed fields; only the inequality* $e_0(S) \leq \|S\|$ *is formalised, which is all that is used anywhere in this blueprint.*

Theorem 10.6. *For every* $S \in \mathcal{L}(X,Y)$*, the sequence* $e_n(S)$ *tends to* 0 *if and only if* $S(B_X)$ *is totally bounded. If* Y *is complete, then*

$$S \text{ is compact} \iff e_n(S) \xrightarrow{n\to\infty} 0,$$

where the forward implication needs no completeness.

[Lean: `SNumbers.isCompactOperator_iff_tendsto_entropyNumber`
`SNumbers.tendsto_entropyNumber_iff_totallyBounded`]

Proof. The first equivalence matches the definitions: a finite ε-net of $S(B_X)$ of size N is a cover by 2^n balls once $2^n \geq N$, and conversely. If S is compact, the closure of $S(B_X)$ is compact, hence totally bounded; if conversely $S(B_X)$ is totally bounded and Y is complete, its closure is compact, which is the definition of a compact operator. □

Chapter 11

Entropy numbers versus s-numbers

Among the s-number sequences, exactly two detect compactness on every Banach space: the Gelfand and the Kolmogorov numbers. The approximation numbers do not — a compact operator into a space without the approximation property is not a limit of finite-rank operators (Enflo) — so the equivalence $a_n(S) \to 0 \iff S$ compact is available only on Hilbert spaces, where all s-numbers agree with a_n.

This chapter compares c_n, d_n and h_n with the entropy numbers of Chapter 10, in both directions: upper bounds for c_n, d_n in terms of e_n give "compact $\Rightarrow$ decay", and a lower bound for e_n gives the sharp comparison. The comparisons are formalised in `EntropyBounds.lean`, the compactness criteria in `AddOns/Compact.lean`, both over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

11.1 Covering estimates

Theorem 11.1 (Covering estimates)**.** *Let $S \in \mathcal{L}(X, Y)$ and suppose every $x \in B_X$ satisfies $\|Sx - y_i\| \le \varepsilon$ for some member of a family $y_1, \ldots, y_k \in Y$. Then*

$$d_k(S) \le \varepsilon \qquad \text{and} \qquad c_k(S) \le 2\varepsilon.$$

[Lean: `SNumbers.kolmogorovNumber_le_of_fin_net`
`SNumbers.gelfandNumber_le_two_mul_of_fin_net`]

Proof. For d_k take $V := \operatorname{span}\{y_1, \ldots, y_k\}$, of dimension at most k. Since $y_i \in V$, the quotient map satisfies $\|\pi_V(Sx)\| = \|\pi_V(Sx - y_i)\| \le \|Sx - y_i\| \le \varepsilon$, so $\|\pi_V \circ S\| \le \varepsilon$.

For c_k choose norming functionals b_i with $\|b_i\| \le 1$ and $b_i(y_i) = \|y_i\|$ (Hahn–Banach) and let M be the common kernel of the $b_i \circ S$, a closed subspace of codimension at most k. For $x \in M$ with $\|x\| \le 1$ pick i with $\|Sx - y_i\| \le \varepsilon$; then $\|y_i\| = b_i(y_i - Sx) \le \varepsilon$ and hence $\|Sx\| \le \|Sx - y_i\| + \|y_i\| \le 2\varepsilon$. ☐

Corollary 11.2 (Dyadic comparison)**.** *For every $S \in \mathcal{L}(X,Y)$ and $n \in \mathbb{N}_0$,*

$$d_{2^n}(S) \;\le\; e_n(S) \qquad \text{and} \qquad c_{2^n}(S) \;\le\; 2\, e_n(S).$$

[Lean: `SNumbers.kolmogorovNumber_two_pow_le_entropyNumber` `SNumbers.gelfandNumber_two_pow_le_two_mul_entropyNumber`]

Proof. An admissible radius ε for $e_n(S)$ provides a net of at most 2^n points, so Theorem 11.1 applies with $k \le 2^n$; both sequences are antitone in the index. ☐

11.2 Compactness criteria

Theorem 11.3. *For $S \in \mathcal{L}(X,Y)$, each of $d_n(S) \to 0$ and $c_n(S) \to 0$ is equivalent to total boundedness of $S(B_X)$. If Y is complete, then*

$$S \text{ is compact} \iff c_n(S) \to 0 \iff d_n(S) \to 0.$$

[Lean: `SNumbers.isCompactOperator_iff_tendsto_kolmogorovNumber` `SNumbers.isCompactOperator_iff_tendsto_gelfandNumber` `SNumbers.tendsto_kolmogorovNumber_iff_totallyBounded` `SNumbers.tendsto_gelfandNumber_iff_totallyBounded`]

Proof. If $S(B_X)$ is totally bounded, a finite ε-net with k points bounds $d_k(S)$ and $c_k(S)$ by Theorem 11.1, and antitonicity gives the decay.

Conversely, suppose $d_n(S) < \varepsilon$ and pick V of dimension at most n with $\|\pi_V \circ S\| < \varepsilon$. Every Sx with $\|x\| \le 1$ is then within ε of a point of V of norm at most $\|S\| + \varepsilon$; that bounded piece of the finite-dimensional space V is compact, so a finite ε-net of it yields a finite 2ε-net of $S(B_X)$.

For $c_n(S) < \varepsilon$ pick a closed M of finite codimension with $\|S|_M\| < \varepsilon$. The image of B_X in the finite-dimensional quotient X/M is bounded, hence totally bounded; choose a finite δ-net inside that image and lift its centres to $u_1, \dots, u_N \in B_X$. If $\|[x] - [u_j]\| < \delta$ there is $m \in M$ with $\|(x - u_j) - m\| < \delta$ and — this is the point — $\|m\| \le 2 + \delta$, a bound *independent of the codimension.* Hence

$$\|Sx - Su_j\| \;\le\; \|Sm\| + \|S\|\,\delta \;<\; \varepsilon(2+\delta) + \|S\|\,\delta,$$

which is small for suitable ε and δ. (Splitting x by a projection with kernel M instead would cost the factor $1+\sqrt{n}$ of Garling–Gordon (Theorem 7.4), and $\sqrt{n}\,c_n(S)$ need not tend to 0.) ☐

Theorem 11.4 (Hilbert spaces: every s-number sequence)**.** *Let H_1, H_2 be Hilbert spaces and s any s-number sequence. Then $S \in \mathcal{L}(H_1, H_2)$ is compact if and only if $s_n(S) \to 0$.*

[Lean: `SVD.isCompactOperator_iff_tendsto_sn`]

Proof. On Hilbert spaces $s_n(S) = a_n(S)$ by the coincidence theorem (Theorem 4.1), and $a_n(S) \to 0$ is equivalent to compactness by Theorem 6.10. ☐

11.3 Entropy bound by Gelfand and Kolmogorov numbers

Proposition 11.5 (Separation bounds e_n from below)**.** *If $S(B_X)$ contains more than 2^n points that are pairwise at distance at least δ, then $e_n(S) \geq \delta/2$.*

[Lean: `SNumbers.le_entropyNumber_of_separated`]

Proof. A ball of radius ε has diameter at most 2ε, so if $2\varepsilon < \delta$ then each ball of an admissible cover contains at most one of the points. Assigning to every point a covering centre is then injective, so there are at least as many centres as points — more than 2^n, contradicting admissibility. ☐

Theorem 11.6 (Triangular flags)**.** *Let $\gamma < d_n(S)$. Then there are $x_0, \dots, x_n \in B_X$ and subspaces $V_0, \dots, V_n \subseteq Y$ with $Sx_k \in V_j$ for $k < j$ and $\|Sx_j + w\| > \gamma$ for all $w \in V_j$. Dually, for $0 < \gamma < c_n(S)$ there are $x_0, \dots, x_n \in B_X$ and functionals $\rho_0, \dots, \rho_n$ with $\|\rho_j\| \leq 1$, $\rho_j(Sx_j) = \gamma$ and $\rho_i(Sx_j) = 0$ for $i < j$.*

[Lean: `SNumbers.exists_kolmogorov_flag` `SNumbers.exists_gelfand_flag`]

Proof. Both are built one vector at a time. For d_n, take V_j to be the span of the images already chosen — of dimension at most n — and pick $x_j \in B_X$ with $\|Sx_j + w\| > \gamma$ for all $w \in V_j$, which is possible because $\gamma < d_n(S)$ and $\dim V_j \leq n$ (Definition 3.7). For c_n, make the same selection below $c_n(S)$ — using $\gamma < c_n(S)$ and Definition 3.5 — on the common kernel of the functionals already chosen, a closed subspace of codimension at most n, and rescale the resulting unit-image vector by γ to land in B_X. ☐

Theorem 11.7 (Pietsch, [7, 12.3.2])**.** *For every $S \in \mathcal{L}(X,Y)$ and $n \in \mathbb{N}_0$,*

$$\max\bigl\{c_n(S), d_n(S)\bigr\} \;\le\; (n+1)\, e_n(S).$$

[Lean: `SNumbers.max_gelfandNumber_kolmogorovNumber_le_succ_mul_entropyNumber`]

Proof. Let $\gamma < d_n(S)$ and take a flag $x_0, \dots, x_n$ as in Theorem 11.6. For $\sigma \in \{\pm 1\}^{n+1}$ set

$$z_\sigma := S\Bigl(\tfrac{1}{n+1}\sum_k \sigma_k x_k\Bigr) \in S(B_X).$$

For $\sigma \neq \tau$ let j be the *largest* index with $\sigma_j \neq \tau_j$. Then $z_\sigma - z_\tau = \frac{1}{n+1}\bigl((\sigma_j - \tau_j)Sx_j + w\bigr)$ with $w \in V_j$, because the indices $k > j$ contribute nothing and those $k < j$ contribute elements of V_j. Since $|\sigma_j - \tau_j| = 2$, the flag property gives $\|z_\sigma - z_\tau\| > 2\gamma/(n+1)$. So 2^{n+1} points of $S(B_X)$ are pairwise more than $2\gamma/(n+1)$ apart, and Proposition 11.5 yields $e_n(S) \ge \gamma/(n+1)$. Letting $\gamma \nearrow d_n(S)$ gives the claim.

For c_n the same computation is read off by the functional ρ_j at the *smallest* disagreeing index j: it annihilates Sx_k for $k > j$, while $\sigma_k = \tau_k$ kills the terms with $k < j$. □

11.4 Entropy bound by Hilbert numbers

Proposition 11.8 (Entropy of the identity)**.** *If $\dim X > n$ then $e_n(\mathrm{id}_X) \ge 1/2$.*

[Lean: `SNumbers.half_le_entropyNumber_id`]

Proof. Let D be the dimension of X as a real vector space, so $D > n$, and let μ be a Haar measure on X. Scaling a ball by ε scales its measure by ε^D, so a cover of B_X by 2^n balls of radius ε gives $\mu(B_X) \le 2^n \varepsilon^D \mu(B_X)$ and hence $1 \le 2^n \varepsilon^D$, the measure of B_X being positive and finite. Were $\varepsilon \le 1/2$, then $2^n \varepsilon^D \le 2^n 2^{-(n+1)} = 1/2$, a contradiction. □

Theorem 11.9 (Hilbert-entropy bound)**.** *For every $S \in \mathcal{L}(X,Y)$ and $n \in \mathbb{N}_0$,*

$$h_n(S) \;\le\; 2\, e_n(S).$$

[Lean: `SNumbers.hilbertNumber_le_two_mul_entropyNumber`
`SNumbers.approximationNumber_le_two_mul_entropyNumber`]

Proof. First let T be an operator between Hilbert spaces and $\gamma < a_n(T)$. The scalar factorisation (Theorem 6.8) provides contractions A, B with $B \circ T \circ A = \gamma \cdot \mathrm{id}$ on ℓ_2^{n+1}; replacing B by $\gamma^{-1}B$ makes the composition the identity, so

$$\tfrac{1}{2} \;\le\; e_n(\mathrm{id}_{\ell_2^{n+1}}) \;\le\; \gamma^{-1} e_n(T)$$

by Proposition 11.8 and the ideal property of e_n. Hence $a_n(T) \le 2e_n(T)$.

Now $h_n(S)$ is the supremum of $a_n(B \circ S \circ A)/(\|B\|\,\|A\|)$ over factorisations through ℓ_2, and each numerator is at most $2e_n(B \circ S \circ A) \le 2\|B\| e_n(S) \|A\|$. □

References

[1] B. Carl, A. Pietsch, *Some contributions to the theory of s-numbers*, Comment. Math. (Prace Mat.) **21**, 65–76, 1978, link.

[2] D. J. H. Garling, Y. Gordon, *Relations between some constants associated with finite dimensional Banach spaces*, Israel J. Math. **9** (1971), 346–361, doi.

[3] F. John, *Extremum problems with inequalities as subsidiary conditions*, in: Studies and Essays presented to R. Courant on his 60th Birthday, Interscience, New York, 1948, 187–204; reprinted in: Fritz John, Collected Papers, Volume 2, Birkhäuser, 1985, 543–560, doi.

[4] M. I. Kadets, M. G. Snobar, *Certain functionals on the Minkowski compactum*, Mat. Zametki **10** (1971), 453–457, link; English transl. in Math. Notes **10** (1971), 694–696, doi.

[5] B. S. Mityagin, G. M. Henkin, *Inequalities between n-diameters*, in: Proc. of the Seminar on Functional Analysis 7, Voronezh, 97–103, 1963.

[6] A. Pietsch, *s-numbers of operators in Banach spaces*, Studia Math. **51** (1974), 201–223, doi.

[7] A. Pietsch, *Operator ideals*, North-Holland Mathematical Library 20, North-Holland, 1980, doi.

[8] A. Pietsch, *Eigenvalues and s-numbers*, Cambridge Studies in Advanced Mathematics 13, Cambridge University Press, 1987, link.

[9] M. Ullrich, *Inequalities between s-numbers*, Adv. Oper. Theory **9** (2024), no. 4, article 82, doi, arxiv.

[10] M. Ullrich, *On bounds between all s-numbers and widths of convex sets*, preprint, 2026, arxiv.